WAVES AND OSCILLATIONS IN NATURE

An Introduction

A. Satya Narayanan
Indian Institute of Astrophysics
Bangalore, India

Swapan K. Saha
Formerly, Indian Institute of Astrophysics
Bangalore, India

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

© 2015 by Taylor & Francis Group, LLC

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20150417

International Standard Book Number-13: 978-1-4665-9094-6 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

© 2015 by Taylor & Francis Group, LLC

Dedicated to
Everyone in pursuit of truth

© 2015 by Taylor & Francis Group, LLC

Contents

List of Figures xv

List of Tables xxi

Principal Symbols xxiii

Preface xxvii

Acknowledgments xxix

1 Introduction to Waves and Oscillations 1
- 1.1 Preamble . . . 2
  - 1.1.1 Electromagnetic Spectrum . . . 3
  - 1.1.2 Types of Spectrum . . . 5
  - 1.1.3 Scattering . . . 7
- 1.2 What Is a Wave? . . . 9
- 1.3 Harmonic Wave . . . 14
  - 1.3.1 Harmonic Plane Waves . . . 14
  - 1.3.2 Harmonic Spherical Waves . . . 16
  - 1.3.3 Fourier Transform Method . . . 18
  - 1.3.4 Phase Velocity . . . 20
  - 1.3.5 Group Velocity . . . 21
  - 1.3.6 Dispersion Relation . . . 23
- 1.4 Monochromatic Fields . . . 25
  - 1.4.1 Complex Representation . . . 25
  - 1.4.2 Superposition Principle . . . 26
  - 1.4.3 Standing Wave . . . 29
  - 1.4.4 Doppler Effect . . . 30
    - 1.4.4.1 Doppler Shift . . . 30
    - 1.4.4.2 Doppler Broadening . . . 31
- 1.5 Intensity of Waves . . . 32
- 1.6 Interference . . . 34
  - 1.6.1 Interference of Two Monochromatic Waves . . . 35
  - 1.6.2 Young's Double-Slit Experiment . . . 39
  - 1.6.3 Michelson Interferometer . . . 41
- 1.7 Diffraction . . . 43

© 2015 by Taylor & Francis Group, LLC

1.7.1 Huygens–Fresnel Principle 43
1.7.2 Kirchhoff's Scalar Diffraction Theory 45
1.7.2.1 Kirchhoff's Diffraction Integral 45
1.7.2.2 Kirchhoff's Boundary Conditions 48
1.7.3 Fresnel–Kirchhoff Diffraction Formula 49
1.7.4 Rayleigh–Sommerfeld Integral 51
1.7.5 Fresnel Approximation 53
1.7.6 Fraunhofer Approximation 55
1.7.6.1 Fraunhofer Diffraction by Square Aperture 56
1.7.6.2 Fraunhofer Diffraction by Slit 59
1.7.6.3 Fraunhofer Diffraction by Circular Aperture 59
1.7.6.4 Point Spread Function 60
1.7.6.5 Resolving Power of a Telescope 61
1.7.6.6 Bessel Function 63
1.7.7 Fraunhofer Diffraction by Grating 64
1.7.7.1 Diffraction by Sinusoidal Amplitude Grating 66
1.7.7.2 Diffraction by Sinusoidal Phase Grating . . 68
1.8 Exercises 69

2 Electromagnetic Waves 73
2.1 Electrostatics 74
2.1.1 Point Charge: Coulomb's Law 74
2.1.2 Electric Energy of a System of Charges 75
2.1.3 Electric Fields 77
2.1.3.1 Field Distributions 77
2.1.3.2 Electric Dipole 78
2.1.3.3 Charge Distributions 79
2.1.4 Electric Flux and Gauss' Law 81
2.1.5 Electric Flux Density 83
2.1.6 Gauss' Law for Electrostatics 85
2.1.7 Potential Difference 90
2.1.7.1 Deriving Electric Field from Electric Potential 91
2.1.7.2 Divergence of the Electric Field 93
2.1.7.3 Poisson and Laplace Equations 94
2.1.7.4 Curl of the Electric Field 95
2.1.8 Field from Line and Surface Charges 96
2.1.9 Current Density 98
2.2 Magnetostatics 99
2.2.1 Magnetic Flux 101
2.2.2 Ampère's Law 102
2.2.3 Biot–Savart Law 102
2.2.4 Magnetic Potential 104
2.2.4.1 Divergence and Curl of **B** 104
2.2.4.2 Magnetic Vector Potential 105

© 2015 by Taylor & Francis Group, LLC

2.2.5 Lorentz Force . . . 106
2.3 Time-Varying Fields . . . 107
2.3.1 Faraday's Law . . . 107
2.3.2 Inductance . . . 108
2.3.2.1 Self-Inductance . . . 108
2.3.2.2 Mutual Inductance . . . 109
2.3.3 Resonant Circuits . . . 110
2.4 Maxwell's Equations . . . 114
2.4.1 Maxwell's First Law . . . 114
2.4.2 Maxwell's Second Law . . . 115
2.4.2.1 Conduction Current Density . . . 115
2.4.2.2 Displacement Current . . . 116
2.4.2.3 Ampère's Law with Maxwell Correction . . . 117
2.4.3 Maxwell's Third Law . . . 118
2.4.4 Maxwell's Fourth Law . . . 119
2.4.5 Maxwell's Equations—Sinusoidal Fields . . . 120
2.4.6 Continuity Equation of Charge . . . 121
2.4.7 Boundary Conditions . . . 122
2.5 Energy Flux of Electrodynamics . . . 124
2.5.1 Poynting Vector . . . 126
2.5.2 Energy Conservation Law of the Electromagnetic Field 127
2.6 Electromagnetic Field Equations . . . 131
2.6.1 General Electromagnetic Wave . . . 133
2.6.2 Harmonic Time Dependence . . . 136
2.6.3 Fourier Transform of Harmonic Equations . . . 139
2.7 Antennas . . . 140
2.7.1 Significance of Antenna Shape . . . 141
2.7.2 Radiation of Electromagnetic Wave by Antenna . . . 142
2.7.2.1 Electric and Magnetic Fields of Oscillating Hertzian Dipole . . . 146
2.7.2.2 Radiation Fields of a Hertzian Dipole Antenna . . . 148
2.7.2.3 Radiation Resistance of an Antenna . . . 148
2.7.2.4 Quarter-Wave Monopole and Half-Wave Dipole . . . 150
2.7.2.5 Power Gain of an Antenna . . . 151
2.7.3 Antennas for Radio Astronomy . . . 151
2.7.4 Waves through Ionosphere . . . 153
2.8 Exercises . . . 156

3 Waves in Uniform Media 159
3.1 Introduction . . . 159
3.2 Simple Harmonic Oscillation . . . 160
3.2.1 Equation for Simple Harmonic Oscillator . . . 162
3.2.2 The Simple Pendulum . . . 163

© 2015 by Taylor & Francis Group, LLC

3.2.3 Angular Simple Harmonic Motion 163
3.2.4 Forced Oscillation and Resonance 163
3.2.5 Resonance 164
3.3 Damped Oscillations 165
3.3.1 Damping by Friction 167
3.4 Coupled Oscillations 169
3.4.1 Superposition of Waves 169
3.4.2 Stationary Waves 169
3.4.3 Coupled Masses 169
3.5 One-Dimensional Wave Equation: D'Alembert's Solution . . 171
3.6 Helmholtz Equation 172
3.7 Normal Mode Eigenvalue Problem 173
3.8 Longitudinal Waves 174
3.9 Traveling Waves 176
3.9.1 Sinusoidal Traveling Waves 176
3.10 Dispersive Waves 177
3.10.1 Linear Evolution Equation 180
3.10.2 Solution of the KdV Equation 180
3.11 Solitons 181
3.11.1 Nonlinear Cubic Schrodinger Equation 184
3.11.2 Two-Soliton Solution 184
3.12 Exercises 185

4 Hydrodynamic Waves **187**
4.1 Introduction 187
4.2 Basic Equations 188
4.2.1 Equations in a Rotating Frame 190
4.3 Small-Amplitude Waves 190
4.3.1 An Application in Geophysics 192
4.4 Gravity Waves 192
4.4.1 Waves in a Steady Stream 195
4.5 Linear Capillary and Gravity Waves 195
4.5.1 One-Dimensional Capillary – Gravity Waves 198
4.6 Surface Waves Generated by a Local Disturbance in the Field 199
4.7 Klein−Gordon Equation 202
4.8 Shallow Water Waves 203
4.8.1 Long Waves in Shallow Water 205
4.9 Boussinesq Equation 208
4.10 Finite Amplitude Shallow Water Waves (Nonlinear Aspects) 210
4.11 Plane Waves in a Layer of Constant Depth 212
4.12 Poincaré and Kelvin Waves 214
4.13 Lamb and Rayleigh Waves 218
4.14 Inertial Waves 221
4.14.1 Axisymmetric Waves 223
4.15 Rossby Waves 223

© 2015 by Taylor & Francis Group, LLC

4.16 Forced Stationary Waves in the Atmosphere 225
4.17 Solitary Waves – KdV Equation 227
4.17.1 KdV Equation 231
4.17.2 Solitary Waves 231
4.18 Exercises 232

5 MHD Waves in Uniform Media 235
5.1 Basic Equations 235
5.2 Sound Waves 238
5.3 Alfven Waves 239
5.4 Shear Alfven Waves 241
5.5 Compressional Alfven Waves 242
5.6 Magneto Acoustic Waves 243
5.7 Internal and Magneto Acoustic Gravity Waves 246
5.7.1 Internal Alfven Gravity Waves 249
5.7.2 Viscous Alfven Gravity Waves 251
5.8 Phase Mixing of Waves 254
5.8.1 Vertical Scale Larger than the Horizontal Scale . . . 258
5.8.2 Uniform Density and Magnetic Field 259
5.9 Resonant Absorption of Waves 260
5.10 Nonlinear Aspects 263
5.11 Exercises 265

6 MHD Waves in Nonuniform Media 267
6.1 Waves at a Magnetic Interface 267
6.1.1 Incompressible Medium 270
6.1.2 Compressible Medium 271
6.2 Surface and Interfacial Waves 273
6.2.1 Special Cases 276
6.2.2 Presence of Steady Flows 277
6.3 Tangential Discontinuity with Inclined Fields and Flows . . 278
6.4 Two-Mode Structure of Alfven Surface Waves 280
6.5 Magneto Acoustic-Gravity Surface Waves with Flows 282
6.6 Waves in a Magnetic Slab 284
6.6.1 Compressible Case 286
6.6.2 Effect of Flows inside the Slab 288
6.6.3 Special Cases 289
6.6.4 Effect of Flows and Gravity with an Application . . . 289
6.7 Negative Energy Waves 292
6.8 Waves in Cylindrical Geometries 293
6.8.1 Different Types of Modes in Cylindrical Geometry . 294
6.9 Slender Flux Tube Equations 297
6.10 Waves in Untwisted and Twisted Tubes 299
6.10.1 Oscillations in Annular Magnetic Cylinders 299
6.10.2 Magnetically Twisted Cylindrical Tube 301

© 2015 by Taylor & Francis Group, LLC

6.11 Applications to Coronal Waves . . . 302
6.11.1 Kink Oscillations . . . 302
6.11.2 Sausage Oscillations . . . 305
6.12 Nonlinear Aspects . . . 306
6.13 Exercises . . . 309

7 Shock Waves 311
7.1 Introduction . . . 311
7.2 Discontinuities in Surfaces . . . 312
7.3 Normal Shock Waves . . . 313
7.4 Oblique Shock Waves . . . 315
7.5 Blast Waves – Similarity Solution of Taylor−Sedov . . . 316
7.6 Weak Shock Waves . . . 319
7.6.1 Example of Weak Shocks . . . 322
7.7 Waves in a Polytropic Gas . . . 323
7.8 An Application of Shock Waves in the Sun . . . 326
7.8.1 An Example from Earth's Bow Shock . . . 328
7.9 Shock Waves in Collisionless Plasmas . . . 328
7.9.1 Dispersive Shock Waves . . . 331
7.10 Shocks in MHD . . . 335
7.10.1 Parallel Shocks . . . 337
7.10.2 Perpendicular Shocks . . . 337
7.11 Nonlinear Studies . . . 341
7.11.1 Burger's Equation . . . 343
7.11.2 Stationary Solutions and Shock Structure . . . 347
7.11.3 Single-Hump Solution . . . 348
7.11.4 Planar N-Wave . . . 349
7.11.5 Backlund Transformation . . . 351
7.12 Exercises . . . 351

8 Waves in Optics 353
8.1 Optical Phenomena . . . 354
8.1.1 Classical Optics . . . 354
8.1.1.1 Geometrical Optics . . . 354
8.1.1.2 Wave Optics . . . 355
8.1.2 Modern Optics . . . 355
8.1.2.1 Quantum Optics . . . 355
8.1.2.2 Statistical Optics . . . 357
8.1.3 Velocity of Light . . . 358
8.2 Nonmonochromatic Fields . . . 359
8.2.1 Complex Representation . . . 360
8.2.2 Power Spectrum . . . 363
8.2.3 Notion of Convolution . . . 363
8.2.3.1 Relationship . . . 364
8.2.3.2 Orthogonality . . . 366

© 2015 by Taylor & Francis Group, LLC

8.2.4 Quasi-Monochromatic Fields . . . 366
8.3 Emission of Wave-Trains . . . 368
8.3.1 Coherence Length and Coherence Time . . . 372
8.3.2 Damped Harmonic Vibrations . . . 375
8.4 Polarization of Plane Monochromatic Waves . . . 377
8.4.1 Instantaneous Optical Field and Polarization Ellipse 379
8.4.2 Stokes Parameters . . . 383
8.4.2.1 Poincaré Sphere . . . 386
8.4.2.2 Degree of Polarization . . . 387
8.4.3 Measurements of Stokes Parameters . . . 389
8.4.3.1 Polarizer . . . 390
8.4.3.2 Retarder . . . 392
8.4.3.3 Rotator . . . 393
8.4.4 Stokes Intensity Formula . . . 394
8.4.5 Jones Matrix . . . 395
8.4.5.1 Jones Matrix for a Polarizer . . . 396
8.4.5.2 Jones Matrix for a Retarder . . . 399
8.4.5.3 Jones Matrix for a Rotator . . . 400
8.4.6 Mueller Matrix . . . 401
8.4.6.1 Mueller Matrix of a Polarizer . . . 402
8.4.6.2 Mueller Matrix of a Retarder . . . 403
8.4.6.3 Mueller Matrix of a Rotator . . . 404
8.4.7 Rotated Polarizing Elements . . . 405
8.4.7.1 Quarter-Wave Retarder . . . 406
8.4.7.2 Half-Wave Retarder . . . 407
8.4.8 Elliptical and Circular Polarizer and Analyzer . . . 407
8.4.8.1 Left-Handed Circular Polarizer . . . 408
8.4.8.2 Left-Handed Circular Analyzer . . . 408
8.4.8.3 Left Circular Analyzer and Polarizer . . . 409
8.4.9 Polarimetry . . . 409
8.4.9.1 Polarimeter . . . 409
8.4.9.2 Astronomical Polarimeter . . . 410
8.5 Exercises . . . 412

9 Plasma Waves 415
9.1 What Is a Plasma? . . . 415
9.2 Plasma Parameters . . . 417
9.2.1 Debye Shielding . . . 417
9.2.2 Plasma frequency . . . 418
9.2.3 Gyro frequency . . . 419
9.2.4 Collision Frequency . . . 419
9.2.5 Plasma as a Dielectric . . . 420
9.3 Electrostatic Waves in Magnetized Plasma . . . 420
9.3.1 Ion Cyclotron Waves . . . 421
9.3.2 Lower Hybrid Frequency . . . 422

© 2015 by Taylor & Francis Group, LLC

9.4 Waves in a Cold Plasma 422
9.4.1 Electromagnetic Waves in a Nonmagnetized Plasma . 422
9.4.2 Application 423
9.4.3 Electromagnetic Waves in a Magnetized Plasma . . . 424
9.4.4 Case 1: Wave Propagation Perpendicular to the Magnetic Field 424
9.4.5 Case 2: Wave Propagation Parallel to the Magnetic Field 426
9.5 Plasma Waves (Warm)−Langmuir Waves 429
9.6 Ion-Acoustic Waves 430
9.7 Waves in Nonhomogeneous Plasmas 432
9.8 Quasi-Linear Theory for Nonhomogeneous Plasmas 438
9.8.1 Quasi-Linear Theory 439
9.9 Nonlinear Waves in Plasmas 443
9.9.1 Pondermotive Force 444
9.9.2 Ion-Acoustic Solitons 445
9.9.3 Nonlinear Schrödinger Equation 447
9.9.4 Zakharaov−Shabat Equation 448
9.10 Exercises 449

10 Fluid and Plasma Instabilities 451
10.1 Introduction 451
10.2 Stability of Parallel Shear Flows 452
10.2.1 Squire's Theorem 454
10.2.2 Rayleigh's Inflexion-Point Theorem 455
10.2.3 Fjortoft's Theorem 455
10.2.4 Howard's Semi-Circle Theorem 456
10.3 Taylor−Goldstein Equation 456
10.4 Orr−Sommerfeld Equation 461
10.5 Rayleigh–Taylor (RT) Instability 464
10.5.1 Application 470
10.6 Kelvin−Helmholtz (KH) Instability 470
10.6.1 Application 474
10.7 Parametric Instability 475
10.7.1 Application 477
10.8 Two-Stream Instability 478
10.9 Interchange (Flute) Instability 482
10.10 Sausage Instability 485
10.11 Kink Instability 486
10.12 Ballooning Instability 487
10.13 Exercises 490

A Typical Tables 493

© 2015 by Taylor & Francis Group, LLC

B Vector Operators **495**
B.1 Vector Formulae 495
B.1.1 Scalar Product 496
B.1.2 Vector Product 497
B.1.3 Del Operator 497
B.1.4 Laplacian Operator 498
B.2 Vector Derivatives 498
B.2.1 Gradient 499
B.2.2 Divergence 499
B.2.3 Laplacian 500
B.2.4 Curl 500
B.3 Stokes Theorem 501
B.4 Green's Theorem 501

Bibliography **503**

Index **513**

© 2015 by Taylor & Francis Group, LLC

List of Figures

1.1 (a) Light dispersion (wave picture) through a prism and (b) dispersion of light quanta (particle picture) through a prism. 3
1.2 Electromagnetic spectrum. 4
1.3 Amplitude pattern. 10
1.4 Plane electromagnetic wave. 15
1.5 Propagation of a plane wave. 16
1.6 Propagation of a spherical wave. 17
1.7 Fourier transform of $sinc\, t$. 18
1.8 Laplace transform of Heaviside unit step. 20
1.9 One-dimensional plot (in time domain) of the cosine factor of $V(z,t) = 2A\cos(\bar{\kappa}z - \bar{\omega}t)\cos(\kappa_g z - \omega_g t)$; the low frequency wave serves as an envelope modulating the high frequency wave. The dashed lines of this curve display the envelope of the resulting wave disturbance. 22
1.10 Interference of the two tilted plane waves gives straight line fringes − the more the tilt the thinner/slimmer the fringes. 38
1.11 2D patterns of $\cos^2(2\pi u x^2)$ fringes when a cylindrical wave interacts with a plane wave. 39
1.12 Illustration of interference with two point sources. 40
1.13 Schematic diagram of a classical Michelson interferometer. . 42
1.14 Construction of the Fresnel zone. 44
1.15 Two closed surfaces around a point, P, where a spherical wavelet has a singularity. 46
1.16 Diffraction by an arbitrary aperture, W. 49
1.17 Contribution from spherical wavelets from an aperture in terms of Huygens−Fresnel principle. 50
1.18 Plane geometry to describe diffraction problem. 52
1.19 (a) Fresnel diffraction curves for $\mathcal{C}$(w), (b) $\mathcal{S}$(w), and (c) the monochromatic Fresnel diffraction patterns. 55
1.20 (a) Sinc function $sinc(x) = \sin(\pi x)/(\pi x)$ (solid line) and the function $J_1(\pi x)/(\pi x)$ (dashed line) and (b) 2D pattern of the Fraunhofer diffraction pattern at a square aperture. 58
1.21 2D pattern of the intensity distribution at a circular aperture. 60
1.22 An example of imaging two point sources at the Rayleigh limit of separation. 62

© 2015 by Taylor & Francis Group, LLC

1.23 Bessel function of the first kind: (a) zero order, $J_0(x)$; (b) first order, $J_1(x)$; and (c) second order, $J_2(x)$. 64
1.24 Amplitude transmittance function of the sinusoidal amplitude grating. 66
1.25 Fraunhofer diffraction pattern for a thin sinusoidal amplitude grating. 67
1.26 Fraunhofer diffraction pattern for a thin sinusoidal phase grating. 69

2.1 (a) The dipole configuration and (b) the electric dipole and calculation of potential at a point P due to the dipole. . . . 78
2.2 Electric flux for non-uniform fields. 81
2.3 Illustration of Gauss' law. 85
2.4 Charged metal sphere. 86
2.5 (a) Using Gauss' law to find the field of a line charge and (b) a line of charge. 87
2.6 Field in a ring of charge. 88
2.7 Field due to a uniform disk of charge. 88
2.8 Infinite plane of charge. 89
2.9 Electric field (a) inside a conductor and (b) inside a container. 90
2.10 Two flat conducting plates parallel to one another separated by a distance, s. 97
2.11 Magnetic field vs current. 100
2.12 Magnetic fields passing through an area. 101
2.13 Field due to a current carrying wire. 102
2.14 Illustration of the Biot−Savart law. 103
2.15 (a) Series resonance circuit and (b) parallel resonance circuit. 112
2.16 Defining divergence and gradient using element of volume ΔV with surface S. 118
2.17 (a) Boundary conditions for the normal components of the electromagnetic field and (b) the integration path in the boundary surface. 122
2.18 Orthogonal triad of vectors. 135
2.19 Antennas for picking up (a) electric and (b) magnetic fields. 141
2.20 (a) Capacitor circuit and (b) dipole showing intrinsic capacitance and charging current. 142
2.21 (a) Electric field, (b) magnetic field, and (c) transverse electromagnetic wave from dipole. 143
2.22 (a) Electrostatic dipole and (b) Hertzian dipole. 144
2.23 (a) Retarded vector magnetic potential and (b) resolved part of vector potential. 145
2.24 Illustration of obtaining the total radiated power. 149
2.25 (a) Dipole and (b) monopole antennas. 150

© 2015 by Taylor & Francis Group, LLC

2.26 (a) A section of radioheliograph at the Gauribidanur Observatory, India, dedicated to observations of the solar corona in the frequency range of 30−150 MHz and (b) one of the parabolic dishes used for the Giant Metrewave Radio Telescope (GMRT) synthesis array at Narayangaon, Pune, India. 152

3.1 Damped harmonic oscillation. 167

4.1 Sketch for a two-dimensional wave. 196
4.2 Dispersive characteristics: C as a function of the wavelength λ. 197
4.3 Simple sketch for a shallow water model. 205
4.4 An example of the symmetric and anti-symmetric Lamb modes. 220
4.5 Two-layer model for a solitary wave. 227
4.6 Simple solution of the KdV equation. 231

5.1 The magnetic field perturbation in the xy plane. 241
5.2 Polar diagram for the Alfven waves. 242
5.3 Polar diagram for the magneto acoustic waves. 243
5.4 Group velocity for the magneto acoustic waves. 244
5.5 Dispersion curves for the three modes for the case $V_A < V_s$. 246
5.6 Dispersion curves for the three modes for the case $V_A > V_s$. 246
5.7 Wave normal surfaces of hydromagnetic-gravity waves: (a) for wave frequencies less than the Brunt–Vaisala frequency ($\omega < 1$) and (b) for wave frequencies greater than the Brunt–Vaisala frequency ($\omega > 1$). 253
5.8 Variation of damping length (solid lines) and wavelength (dashed lines) as a function of normalized frequency at different values of a normalized wavenumber: (a) for $\omega < 1$ and (b) for $\omega > 1$. 254
5.9 Phase mixing of Alfven waves. 256
5.10 The inhomogeneous magnetic region with a continuously varying Alfven speed. 260
5.11 Doppler shifted Alfven frequency as a function of the radius. 262
5.12 Same as Figure 5.11 for a different density profile. 262

6.1 Possible regions for surface wave propagation for specific values of the interface parameters. 273
6.2 Dispersive characteristics of surface waves with flows, for specific parametric values. 279
6.3 Same as in the Figure 6.2 for different parametric values. . 280
6.4 Solution of the dispersion relation for MAG with flow for specific values. 282
6.5 Solution of the dispersion relation for MAG with flow for specific values. 283

© 2015 by Taylor & Francis Group, LLC

6.6 Different types of modes possible in a flux tube. 285
6.7 Magnetic field in a structured slab. 286
6.8 The phase-speed ω/k_z, plotted as a function of the nondimensional wavenumber $k_z x_0$. 287
6.9 The geometry. 290
6.10 The normalized phase velocity as a function of the nondimensional wavenumber. 291
6.11 Cylindrical geometry with uniform flow inside the tube. . . 293
6.12 Solution of the dispersion relation for the cylindrical geometry. 297
6.13 Distribution of the magnetic field in the annulus. 300
6.14 Distribution of the magnetic field in the twisted tube. 301
6.15 A loop oscillation. 303
6.16 Behavior of $F_m(ka)$ for different values of ka. 303
6.17 Variation of the magnetic field for different coronal parameters. 304
6.18 Dependence of the cut-off wavenumber of the sausage mode. 305

7.1 Hugoniot curve. 315
7.2 Velocity, pressure, and density (nondimensional) as a function of scaled radius behind the Sedov−Taylor blast wave. 318
7.3 Left: for the linear hyperbolic system, the solution is a simple superposition of traveling waves. Right: for the nonlinear system, waves of different families have nontrivial interactions. 320
7.4 After a time, T, when the characteristics start to intersect, a shock is produced. 321
7.5 Ratio of the pressure as a function of V_2/V_1. 324
7.6 An example of collisionless shock produced by the interaction of the solar wind and the Earth's magnetosphere. 329
7.7 Types of dispersion curves. 332
7.8 Dispersive shocks with trailing ($a > 0$) and leading ($a < 0$) wave train. 333
7.9 Schematic view of a bow shock − The formation of parallel and perpendicular shocks is also illustrated. 337

8.1 Hilbert transform of a rectangular as well as a sinusoidal vibration. 362
8.2 Left panel: a sinusoidal function, and right panel: its power spectrum. 363
8.3 Convolution of two truncated exponentials. 364
8.4 Fourier transform of a sinusoidal vibration. 367
8.5 Top panel left: finite sinusoidal wave-train and right: the corresponding spectrum. Bottom panel left: large number of sinusoidal wave-trains and right: the corresponding spectrum. 370

© 2015 by Taylor & Francis Group, LLC

8.6 The function $[\sin \pi(\nu - \nu_0)\tau_c/\pi(\nu - \nu_0)\tau_c]^2$. 374
8.7 Left panel: complex amplitudes of a large numbers of decaying wave-trains and right panel: variation of $\mid \widehat{U}^{(r)}(\nu) \mid^2$ as a function of ν. 376
8.8 Description of polarization ellipse (a) in terms of $x,\ y$ and (b) in terms of $x',\ y'$ coordinates. 380
8.9 Lissajous representations of polarization ellipses for various values of the phase difference δ. 382
8.10 Poincaré representation of polarized light on a sphere. . . . 387
8.11 Schematic representation for optical element aligned with axes (x, y). 392
8.12 Rotation of optical field components by an optical element through angle ϕ, about γ axis; the optical element is rotated by the angle ϕ, about the z axis, so that $0 \leq \phi < \pi$. 394
8.13 Three-element, circular polarizer formed from two quarter-wave plates `A`, `C` and a linear polarizer, `B`. Angle of linear polarizer; $\phi = 45°$ for LCP and $\phi = -45°$ for RCP. 400

9.1 Solutions for the (a) R and L waves (parallel magnetic field), and (b) O and X waves (perpendicular magnetic field). . . . 425
9.2 Parallel propagation for the (a) high plasma density and (b) low plasma density. Cutoffs are found when the dispersion curve meets the frequency axis ($k = 0$) and resonances are found for large values of k. 428
9.3 Solution of the dispersion relation for a right-handed wave propagating parallel to the magnetic field in a magnetized plasma. 430
9.4 The electrostatic Langmuir wave (electron plasma wave) is the horizontal dot-dashed line at $\omega/\omega_{pe} = 1$. 431

10.1 Three-layered model due to Gossard. 458
10.2 Neutral curves:solid for the present model; dashed for the three-layer model Gossard's. 460
10.3 Streamline pattern of this model. 460
10.4 A simple sketch of the shear layer. 462
10.5 The solution for the frictionless case. 463
10.6 The solution with the inclusion of friction. 464
10.7 A simple sketch of the Rayleigh–Taylor configuration. . . . 467
10.8 A simple sketch for the Kelvin−Helmholtz instability. 471
10.9 Plots of the three functions for determining the stability criterion. 480
10.10 Solutions of Equation (10.157) plotted with x as a function of α. 481
10.11 Ion and electron drifts and the resultant electric field for interchange instability. 483

© 2015 by Taylor & Francis Group, LLC

10.12 A diagram for sausage instability. 485
10.13 A simple sketch for kink instability. 487
10.14 The maximum stable pressure gradient α as a function of the shear parameter S of the ballooning mode. 489

© 2015 by Taylor & Francis Group, LLC

List of Tables

1.1 Locations of the First Five Maxima of the Function 58
1.2 Locations of the Maxima and Minima of the Airy Pattern . 61

5.1 MHD Waves 245

9.1 Different Types of Plasma Waves 433

A.1 Comparison between Gravitation and Electrostatics 493
A.2 Static Electric Field Relations 493
A.3 Static Magnetic Field Relations 494
A.4 Maxwell's Equations 494

© 2015 by Taylor & Francis Group, LLC

Principal Symbols

$\mathbf{a}$	Acceleration
A	Ampere
A	Complex amplitude of the vibration
$\mathbf{A}$	Vector potential
$\mathbf{B}$	Magnetic induction
$\mathbf{B}(\mathbf{r}, t)$	Time dependent magnetic field
$\mathbf{B}_0$	Amplitude of the magnetic field vector
C	Capacitance
C	Coulomb
c	Velocity of light
D	Diameter of the aperture
$\mathbf{D}$	Electric flux density
$\mathbf{E}$	Electric field vector
$\mathbf{E}(\mathbf{r}, t)$	Time dependent electric field
$\mathbf{E}_0$	Amplitude of the electric field vector
$\mathcal{E}$	Electromotive force
$\mathbf{F}$	Force
F	Farad
f	Electrical frequency
G	Gain
$\mathbf{H}$	Magnetic field
H	Henry
$\hbar (= h/2\pi)$	Reduced Planck constant
i	Current
I	Intensity of light
$\Im$	Imaginary part of the quantity
J	Jones matrix
$\mathbf{J}$	Current density
$\mathbf{J}_c$	Conduction current density
$\mathbf{J}_v$	Convection current density
j	= 1, 2, 3 (indices)
J	Joules
$J_0(x)$	Bessel function of the first kind and order zero
$J_1(x)$	Bessel function of the first kind and order one
L	Inductance
l_c	Coherence length
M	Mueller matrix
$\mathbf{M}$	Magnetization
m	Mass

© 2015 by Taylor & Francis Group, LLC

m_0	Rest mass
N	Integer value
N	Newton
n	Refractive index
$\hat{\mathbf{n}}$	Unit vector
$\mathbf{p}$	Momentum
P	Pressure
$\mathcal{P}$	Polarization
q	Electron charge
$\mathbf{r}(= x, y, z)$	Position vector of a point in space
R	Resistance
$\Re$	Real part of the quantity
R_L	Load resistance
S	Stokes parameter
$\hat{\mathbf{s}}$	Unit vector
$S(\mathbf{x})$	Point spread function
$\mathtt{S}$	Surface
$\mathbf{S}(\mathbf{r}, t)$	Poynting vector
$\mathcal{S}_r$	Strehl's ratio
sr	Steradian
t	Time
T	Period
$U(\mathbf{r}, t)$	Complex representation of the analytical signal
$\mathbf{u} = (u, v)$	Spatial frequency vector
V	Electrostatic potential
$V(\mathbf{r}, t)$	Monochromatic optical wave
v	Velocity
$\mathcal{V}$	Visibility
$\mathtt{V}$	Volume
v_g	Group velocity
v_p	Phase velocity
W	Work
$\mathbf{x}(= x, y)$	Two-dimensional (2-D) position vector
*	Complex operator
$\star$	Convolution operator
$\langle\ \rangle$	Ensemble average
$\widehat{\ }$	Fourier transform operator

© 2015 by Taylor & Francis Group, LLC

Greek Symbols

β	Orientation angle
γ	Lorentz factor
δ	Phase difference
$\delta(x)$	Dirac delta function
ν	Optical frequency
$\Delta\nu$	Spectral width (Optical)
Δf	Electrical bandwidth
Δt	Integration time
$\Delta\varphi$	Optical path difference
ϵ	Permittivity or dielectric
ϵ_0	Permittivity in vacuum
θ	Angular diameter
(θ, ϕ)	Polar coordinates
κ	Wavenumber
$\boldsymbol{\kappa}$	Wave vector
λ	Wavelength
λ_0	Wavelength in vacuum
μ	Permeability of the medium
μ_0	Permeability in vacuum
ν	Frequency
$\boldsymbol{\xi}(\xi, \eta)$	2-D position vector
ρ	Charge density
σ	Specific conductivity
τ_c	Coherence time
ϕ	Rotating angle
Φ_B	Magnetic flux
Φ_E	Electric flux
φ	Electrostatic potential
χ	Ellipticity angle
$\psi(\mathbf{r}, \nu)$	Phase function
Ψ	Time-dependent wave-function
Ω	Solid angle
ω	Angular frequency
$\boldsymbol{\nabla}$	Linear vector differential operator
$\boldsymbol{\nabla}^2$	Laplacian operator

© 2015 by Taylor & Francis Group, LLC

Some Numerical Values of Physical and Astronomical Constants

c	Speed of light in free space	$3 \times 10^8\,\mathrm{m.s^{-1}}$
eV	Electron volt	$1.60 \times 10^{-19}\ J$
G	Gravitational constant	$6.672 \times 10^{-11}\,\mathrm{kg^{-1}.m^3.s^{-2}}$
h	Planck's constant	$6.626096 \times 10^{-34}\mathrm{J.s}$
q	Elementary charge	1.6×10^{-19} C
ϵ_0	Permittivity constant	$8.8541 \times 10^{-12}\,\mathrm{F.m^{-1}}$
μ_0	Permeability constant	$1.26 \times 10^{-6}\,\mathrm{H.m^{-1}}$

List of Acronyms

AC	Alternating current
BC	Babinet compensator
CCD	Charge coupled device
dB	Decibel
DC	Direct current
EMF	Electromotive force
EUV	Extreme ultraviolet
FT	Fourier transform
FWHM	Full width at half maximum
HF	High frequency
Hz	Hertz
IF	Intermediate frequency
IR	Infrared
keV	Kilo electron-volt
Laser	Light amplification by stimulated emission of radiation
LCP	Left-handed circular polarizer
LED	Light emitting diode
LF	Low frequency
LHS	Left-hand side
LO	Local oscillator
MHz	Megahertz
MMF	Magnetomotive force
nm	Nanometer
NUV	Near UV
OPD	Optical path difference
PSF	Point spread function
RCP	Right-handed circular polarizer
RHS	Right-hand side
RMS	Root-mean-square
UV	Ultraviolet

© 2015 by Taylor & Francis Group, LLC

Preface

Waves and oscillations are found everywhere in nature. They are present at large scales (galactic), as well as at microscopic scales (neutrino). The dynamics and behavior of these oscillations depend a lot on the nature of the medium through which they propagate. For example, we have hydrodynamic waves (water waves), shock waves in a medium which is compressible. We also come across MHD (magneto hydrodynamic) waves, present in a magnetized medium. When the matter is ionized, one encounters plasma waves and so on.

The aim of this book is to present, at an introductory level, the different types of waves and oscillations that one observes and studies, from macroscopic to microscopic scales. The first chapter deals with an introduction to electromagnetism, the different types of electromagnetic spectra, wave and its characteristics such as phase velocity, group velocity and dispersion relation, applicable to all types of waves. Emphasis on application to astrophysics is introduced at this stage for setting up the theory with possible applications in astrophysics. The notion of interference, diffraction, and their importance in observational astronomy is mentioned briefly. The second chapter is a discussion on Maxwell's equations, used to study electromagnetic waves.

An introduction to waves in a uniform medium is presented in Chapter 3. Topics such as simple harmonic motion, simple pendulum, forced and free oscillations, along with resonance are developed from first principles. Damped and coupled oscillators are discussed briefly. The mathematical description of these waves is introduced and the corresponding solutions are derived briefly. The notions of normal mode eigenvalue problem and dispersion of waves are discussed in subsequent sections. A brief discussion on solitons is introduced at the end of the chapter. The concept of waves and oscillations in hydrodynamics is introduced in Chapter 4. The basic equations in rotating and nonrotating fluids are discussed. The effect of gravity, stratification, Coriolis force, and long wavelength approximation are studied for small and finite amplitude waves. Nonlinear aspects, which lead to the KdV equation (solitary waves), are discussed in brief.

Waves in a magnetized medium, for both homogeneous and nonhomogeneous media, are discussed in Chapters 5 and 6. The linearized equations of MHD waves are derived from Maxwell's equations and the dispersion relation for the Alfven, fast and slow magnetoacoustic waves, is derived from first principles. Effects of uniform flow, gravity, and density stratification are dealt with briefly. Application to solar physics and nonlinear aspects including finite amplitude effects are discussed briefly. The notion of resonant absorp-

© 2015 by Taylor & Francis Group, LLC

tion and phase mixing as possible mechanisms for coronal heating are also introduced.

Different types of shock waves, such as normal and oblique shocks, are studied in Chapter 7. The concept of blast wave and the solution from Taylor and Sedov is introduced in this chapter. Weakly nonlinear aspects of shock waves are also discussed briefly. An application of shock waves in the sun is presented here. Shocks in MHD and collisionless plasmas are also dealt with briefly. Nonlinear studies, which include an introduction of Burger's equation, stationary, and single-hump solutions, are also presented. The notion of the Planar N-wave and the Backlund transformation is found at the end of the chapter. Chapter 8 deals with waves in optics. The notion of classical and modern optics is presented briefly. Nonmonochromatic fields and their properties are discussed at length. One of the important concepts of waves in optics is the notion of polarization. A discussion on Stokes parameters and their measurements is presented. Topics which deal with polarization from experimental and observational points of view are presented in detail. For example, the notions of polarizer, retarder, and rotator are discussed briefly.

Chapter 9 deals with waves and oscillations in plasmas (the fourth state of matter). It starts with the basic definition of a plasma, the different plasma parameters such as Debye shielding, plasma frequency, gyro frequency, and collision frequency, etc. A discussion on electrostatic waves in a cold as well as in a normal plasma is presented from first principles. The effect of an magnetic field, which transforms an electrostatic wave to an electromagnetic wave, is presented. Langmuir waves (warm plasma) and ion-acoustic waves which arise due to compressibility and charge effects are discussed in brief. Quasi-linear as well as nonlinear aspects of plasma waves are presented at the end of the chapter. For example, the nonlinear Schrodinger equation and the Zakharaov−Shabat equation and their solutions are presented briefly. The final chapter deals with the notion of instabilities in hydrodynamics and plasmas. Some of the important instabilities, such as the Kelvin−Helmholtz instability, Taylor−Goldstein instability, and parametric instabilities, are presented in this chapter. Also plasma instabilities such as the two-stream, sausage, kink, and ballooning are presented at the end of the chapter.

We have provided some simple exercises at the end of each chapter. This will enable the student to have a good grasp of the basics involved in each of the topics covered. The book is far from complete. In fact, books and monographs have been written in the past for each of the types of the waves presented in this book. These works are more technical and exhaustive. One would agree that a work which covers several aspects of waves and oscillations is a very ambitious project and it would be impossible to do a good justification in bringing out such a work. However, this is an earnest effort by us to write a book at an introductory level on the various types of waves and oscillations that one encounters in nature. The long list of bibliography sources will enable the interested reader to get into more technical aspects. We hope that this book will be a welcome introduction to researchers working in different areas of physics and hopefully serve as a good reference book.

© 2015 by Taylor & Francis Group, LLC

Acknowledgments

A. Satya Narayanan would like to express his gratitude to his PhD supervisor, the late Professor P. L. Sachdev, who introduced him to the exciting field of waves and oscillations. He had the privilege of learning under him the various techniques for solving linear and nonlinear waves. He also had the privilege of attending the lectures given by Professor G. B. Whitham on shock waves and solitary waves. He is grateful to Dr. K. Somasundaram for introducing him to the exciting area of waves in magneto hydrodynamics (MHD). There are more stalwarts who have contributed to the area of MHD waves, to name a few, B. Roberts, R. Erdelyi, M. S. Ruderman, M. Goossens, V. M. Nakariakov, and T. Sakurai. Apologies to those whose names are not mentioned. The cover of this book is based on the figures of papers by B. Roberts, R. Erdelyi, and V. M. Nakariakov. He is grateful to them for permitting him to use their work in the cover design.

Professor Narayanan is grateful to his colleague, friend, and critic R. Ramesh for going through the manuscript critically, providing suggestions and improvements. Rajat K. Chaudhuri is thanked for giving confidence that the author could do it. C. Kathiravan helped in the cover design. Mr. Sarfraz Khan from Taylor & Francis is thanked for his help in designing the cover. He is happy to thank Dr. Sreepat Jain for encouragement and advice. Dr. Aastha Sharma of Taylor & Francis Group is thanked for her help and correspondence. Her quick response and guidance helped a lot in the writing process. CRC Press personnel have done a wonderful job and the authors are grateful to them. He thanks his family, Sukanya and Prahladh S. Iyer for their patience and support. Dr. P. Sreekumar, the director of Indian Institute of Astrophysics, is thanked for his permission and the support extended to the authors.

Swapan K. Saha acknowledges the support and help from the following people: A. Basuray, B. A. Vargese, S. K. Sithoria, A. Surya, A. S. Somkuwar, and K. Goswami.

A. Satya Narayanan
Swapan K. Saha (Retired)
Indian Institute of Astrophysics, Bangalore, India

© 2015 by Taylor & Francis Group, LLC

Chapter 1

Introduction to Waves and Oscillations

1.1 Preamble 2
1.1.1 Electromagnetic Spectrum 3
1.1.2 Types of Spectrum 5
1.1.3 Scattering 7
1.2 What Is a Wave? 9
1.3 Harmonic Wave 14
1.3.1 Harmonic Plane Waves 14
1.3.2 Harmonic Spherical Waves 16
1.3.3 Fourier Transform Method 18
1.3.4 Phase Velocity 20
1.3.5 Group Velocity 21
1.3.6 Dispersion Relation 23
1.4 Monochromatic Fields 25
1.4.1 Complex Representation 25
1.4.2 Superposition Principle 26
1.4.3 Standing Wave 29
1.4.4 Doppler Effect 30
1.4.4.1 Doppler Shift 30
1.4.4.2 Doppler Broadening 31
1.5 Intensity of Waves 32
1.6 Interference 34
1.6.1 Interference of Two Monochromatic Waves 35
1.6.2 Young's Double-Slit Experiment 39
1.6.3 Michelson Interferometer 41
1.7 Diffraction 43
1.7.1 Huygens–Fresnel Principle 43
1.7.2 Kirchhoff's Scalar Diffraction Theory 45
1.7.2.1 Kirchhoff's Diffraction Integral 45
1.7.2.2 Kirchhoff's Boundary Conditions 48
1.7.3 Fresnel–Kirchhoff Diffraction Formula 49
1.7.4 Rayleigh–Sommerfeld Integral 51
1.7.5 Fresnel Approximation 53
1.7.6 Fraunhofer Approximation 55
1.7.6.1 Fraunhofer Diffraction by Square Aperture 56
1.7.6.2 Fraunhofer Diffraction by Slit 59
1.7.6.3 Fraunhofer Diffraction by Circular Aperture 59
1.7.6.4 Point Spread Function 60
1.7.6.5 Resolving Power of a Telescope 61
1.7.6.6 Bessel Function 63
1.7.7 Fraunhofer Diffraction by Grating 64
1.7.7.1 Diffraction by Sinusoidal Amplitude Grating 66
1.7.7.2 Diffraction by Sinusoidal Phase Grating 68
1.8 Exercises 69

© 2015 by Taylor & Francis Group, LLC

1.1 Preamble

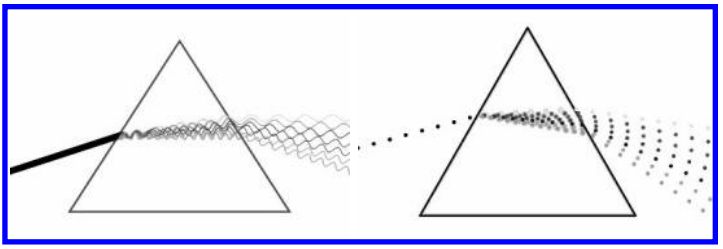
Development of the mathematical theory by James Clerk Maxwell (1831−1879) was a major breakthrough[147], which led to the discovery of the electromagnetic radiation. An exact description of electromagnetic radiation and its behavior requires a knowledge of Maxwell's electromagnetic field equations that describe the temporal and spatial dependence of electromagnetic fields and provide good agreement with observed phenomena over a wide range of frequencies in the spectrum. The spectrum is the variation of the intensity of the radiation as a function of the frequency or wavelength. Around 1864, Maxwell predicted the existence of electromagnetic waves that travel through space at the speed of light. Heinrich Rudolf Hertz (1847−1894) proved Maxwell's laws in the late 1880s when he produced electromagnetic waves using an oscillating current generated by the spark of an induction coil[25].

Electromagnetic waves, also called the electromagnetic radiation, are produced by the motion of electrically charged particles. The radiation is emitted or absorbed when an atom or a molecule moves from one energy level to another. An individual atom consists of a nucleus made up of a core of protons (positively charged particles) and neutrons (particles having no charge) surrounded by electrons (negatively charged particles). The number of electrons and protons is equal, such that the atom is overall electrically neutral. The electrons occupy certain energy levels, based on the number of electrons in the atom, which is different for each element in the periodic table. The charge of an electron is $-e$ and that of a proton is $+e$. The charge of any macroscopic system is an integral multiple of e, where $e = 1.602 \times 10^{-19}$ C, in which C stands for coulomb.

When an electric charge vibrates, the electric field around it changes creating a changing magnetic field. All matter contains charged particles that are always moving; hence, all objects emit electromagnetic waves; the wavelengths become shorter as the temperature of the material increases. Besides acting like waves, it behaves as a stream of particles called the photons. A photon is a quantum or packet of electromagnetic radiation whose energy depends on the frequency of the waves. The energy of a photon varies only with the frequency of the photon, regulated by the speed of light which is a constant; the photons with the highest energy correspond to the shortest wavelengths. A photon in free space travels at the speed of light, which is 299,792,458 meters per second ($m.s^{-1}$). Photons have properties like energy $\hbar\omega$, zero rest mass, momentum, and spin $\hbar/2$; $\hbar$ is Planck's constant, named after Max Karl Ernst Ludwig Planck (1858−1947), $h = 6.626196 \times 10^{34}$Joules (J), divided by 2π. Photons are noninteracting bosons.

In fact, the birth of quantum mechanics is intimately linked with the theories and discoveries relating to the nature of light. Sir Isaac Newton

© 2015 by Taylor & Francis Group, LLC



(a) (b)

FIGURE 1.1: (a) Light dispersion (wave picture) through a prism and (b) dispersion of light quanta (particle picture) through a prism.

(1642−1727) described light as a stream of particles, what was termed as corpuscular (little particles) theory. This view was superseded by the wave picture of light developed by Christian Huygens (1629−1695), known as the Huygens wavefront theory (see Section 1.7), a picture that culminated in the electromagnetic theory of Maxwell. The wave−corpuscle controversy was overcome in the quantum theory of light, which has particle and wave aspects depending on the phenomena considered. Figure (1.1) depicts the concepts of both wave picture and particle picture.

1.1.1 Electromagnetic Spectrum

The electromagnetic spectrum has a continuous energy spectrum, a graph that depicts the intensity of light being emitted over a range of energies (see Figure 1.2). However, the entire distribution covers a wide range of frequencies and wavelengths, and it consists of many subranges. The human eye responds to visible light, whereas detecting the rest of the spectrum requires various types of instruments; in ultraviolet (UV) and shorter wavelengths, observations are carried out from space. The electromagnetic spectrum is divided into various regions extending for different frequency or wavelength intervals.

1. Gamma rays: The frequency of a gamma ray is $> 3 \times 10^{19}$ hertz (Hz; cycles per second), and its wavelength is $\lambda < 1$ nm (nanometer). Its typical energy at 300 exahertz (EHz $= 10^{18}$ Hz) is 1.24 megaelectron-volts (MeV); the unit of energy is the electron-volt[1] (eV). These radiations are associated with cosmic sources. The other sources are the gamma decay of radioactive materials and nuclear fission.

[1]The unit of energy is the electron-volt (eV), where 1 eV is the amount of kinetic energy acquired by a single unbound electron accelerated through an electric potential difference of 1 volt (1joule per coulomb; 1 J.C^{-1}), that is,

$$1\ eV = 1.602176565(35) \times 10^{-19}\ \text{J},$$

$e(= 1.60217665(35) \times 10^{-19}$ C) is the elementary charge, and V is the voltage.

© 2015 by Taylor & Francis Group, LLC

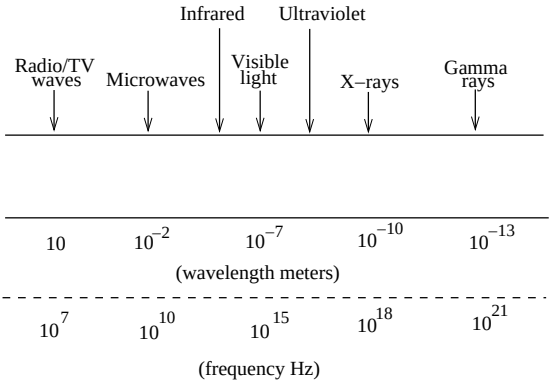


FIGURE 1.2: Electromagnetic spectrum.

2. X-rays: The frequency range for X-ray lies between 3×10^{16} to 3×10^{19} Hz (which corresponds to a range of photon energies from about 120 eV to 100 keV (million electron volts). The X-ray radiation is divided into

   (a) Hard X-rays fall between 3 and 30 EHz, corresponding to energy of 12.4 kiloelectron-volt (keV) to 124 keV.

   (b) Soft X-rays that lie between 300 petahertz (PHz = 10^{15}) to 3 EHz, with corresponding energy of 124 eV and 12.4 keV.

   These radiations are commonly produced by accelerating (or decelerating) charged particles.

3. Ultraviolet (UV) rays: The frequencies in UV rays range between 7×10^{14} to 3×10^{16} Hz (whose energy falls between 3 and 120 eV). The UV radiation can be categorized as

   (a) Extreme-UV (EUV) radiation, whose frequencies line between 1.5 to 30 PHz; the wavelengths are λ = 10-200 nm.

   (b) Near-UV (NUV) radiation, the frequencies of which are 0.7−1.5 PHz and wavelengths are λ = 200-400 nm.

   Unlike X-rays, ultraviolet radiation has a low power of penetration.

4. Optical band: The frequencies of optical band (visible band) fall between 4×10^{14} and 7×10^{14} Hz (λ 400-700 nm). The visible part may be further subdivided according to color, for instance, violet, indigo, blue, green, yellow, orange, and red, with red at the long wavelength end and violet at the short wavelength end.

5. Infrared rays: The infrared frequencies are 3×10^{11} to 4×10^{14} Hz; heat radiation is the source for infrared frequencies. IR radiation extends

© 2015 by Taylor & Francis Group, LLC

from the red edge of the visible spectrum at ∼700 nm to 1 mm (beyond which the microwave region commences). The IR, in general, is divided into five sub regions[110], for instance,

(a) near-IR (0.7−1.4 μm),
(b) short-wavelength IR (SWIR; 1.4−3.0 μm),
(c) mid-wavelength IR (MWIR; 3.0−8.0 μm),
(d) long-wavelength IR (LWIR; 8.0−14.0 μm), and
(e) very long-wavelength IR (VLWIR; 20.0−1,000.0 μm).

IR radiation possesses low energy.

6. Microwave: Microwave radiation has a frequency within the range of 3×10^9 Hz to 3×10^{11} Hz and a wavelength between 1 mm and 10 cm (centimeter).

7. Radio waves: The radio frequency spectrum is divided into several bands, for example,

(a) very high frequencies (VHF; 30−300 MHz, λ 10−1 m),
(b) high frequencies (HF; 3−30 MHz, λ 100−10 m),
(c) medium frequencies (MF; 300−3000 kHz, λ 1,000−100 m),
(d) low frequencies (LF; 30−300 kHz, λ 10,000−1,000 m), and
(e) very low frequencies (VLF; 3−30 kHz, λ 100,000−10,000 m).

The Earth's atmosphere blocks off the incoming radiation from space beyond the violet edge of the visible spectrum, while it attenuates the infrared (IR). The detection can be made from the ground through the atmospheric windows, which allow the radiations in the visible, IR, and radio wavelengths to penetrate the atmosphere. The human eye responds to visible light, while detecting the rest of the spectrum requires various types of instruments; in ultraviolet (UV) and shorter wavelengths, observations are carried out from space.

1.1.2 Types of Spectrum

The percentage of absorbing atoms or ions is connected with the spectral lines, which provide the abundance of the atom producing the line and can be estimated by quantum mechanics and a model of atmospheric line formation theory. Spectrum is mainly of two types:

1. Continuous spectrum: Continuous spectrum, also called the thermal spectrum, possesses energy at all wavelengths or colors, for instance, a rainbow. Most continuous spectra are from hot, dense objects. In the emerging spectrum of a star, bound-free and free-free processes give rise to continuum spectrum and scattering modifies the continuum.

© 2015 by Taylor & Francis Group, LLC

2. Discrete spectra: Discrete spectra that appear like a bunch of lines possesses energy at certain wavelengths. These lines are produced by gases (of either atoms or molecules). A discrete spectrum is more complex because it depends on temperature, chemical composition of the object, the gas density, surface gravity, and speed.

   (a) Absorption spectra: An atom consists of a nucleus surrounded by electrons that swarm around it. These electrons stay in tracks, called the energy levels, at specific distances from the nucleus, which jump from an inner level to an outer level. This makes an absorption line. When light with a continuous spectrum is passed through a relatively cool gas, the wavelengths characteristic of the gas are absorbed; hence the spectra appear as dark lines, called the Fraunhofer lines, superimposed on a continuous background. The bound-bound transitions give rise to line spectrum. The absorption lines arise when the atoms, which are in a lower state, absorb energy and get excited to higher energy levels.

   (b) Emission spectra: Unlike with absorption line, if the electron jumps from an outer level to an inner level, it gives out a photon. This makes an emission line. Hot gas under low pressure produces an emission line, as well as continuum spectrum. The lines are brighter than the background spectrum which could provide information about flares and other atmospheric activity in the star. These lines are produced if the atoms return from the higher energy states to a lower state by emitting radiations. Line strengths of selected forbidden emission lines could provide the temperature, while intensity ratio of two suitable pairs of emission lines can also yield density of plasma. The Doppler-shift in the line profile tells about the line-of-sight velocity as well.

   Studies of both absorption and emission lines in a spectrum of a star are important because these are the fingerprints of the atoms and molecules that make up the star; spectroscopic studies can reveal these aspects.

Spectroscopy is the study of the interaction between electromagnetic radiation and matter. It endeavors to measure the reflectance or transmission of a sample as a function of the wavelength. It analyzes the lines of light emitted from excited atoms as the electrons drop back through their orbitals. This method can be employed to understand the characteristics, such as, velocities, red-shifts, abundances, and magnetic field. It works on the principle that, under certain conditions, materials absorb or emit energy, which can be adapted in several ways to extract the information, namely, energies of electronic, vibrational, rotational states, structure and symmetry of molecules, dynamic information. A spectrograph (spectrometer) is an instrument that collects, spectrally disperses, and re-images a signal. It is designed to measure the distribution of radiation of a source in a particular wavelength re-

© 2015 by Taylor & Francis Group, LLC

gion. The output signal is a series of monochromatic images corresponding to wavelengths present in the light imaged at the entrance slit (see Section 1.7.6.2). Its principal components are a monochromator and a detector, such as CCD (charge-coupled device). The radiant power enters the entrance slit of the monochromator that selects a narrow spectral band of radiant power and transmits it through the exit slit to the photosensitive surface of the detector. A spectrometer requires the following components:

- a collimator (a system of lenses, which renders a beam of rays parallel),
- a dispersing element, namely, a grating (see Section 1.7.6.3), which spreads the light intensity in space as a function of wavelength, and
- a focusing element by which the image is formed at the exit slit of a monochromator and at the detector focal plane.

1.1.3 Scattering

Scattering is a physical process in which a beam of particles or radiation disperses into a range of directions as a result of physical interactions. This term is also used for the diffusion of electromagnetic waves by the atmosphere. The scattering medium introduces random path fluctuations on reflection or transmission. The large and rapid variations of the phases are the product of these fluctuations with wave vector, $\boldsymbol{\kappa}$. A change in frequency of the light changes scale of the phase fluctuations. There are two broad types of scattering:

1. Elastic scattering: In elastic scattering the photon energies of the scattered photons are unaltered with hardly any loss or gain of energy by the radiation. Based on a dimensionless size parameter, $\beta = \pi D_p/\lambda$, in which πD_p is the circumference of a particle and λ the wavelength of incident radiation, the models of light scattering are divided into the following domains.

   (a) Rayleigh scattering: Named after Lord Rayleigh (1842−1919), Rayleigh scattering is caused by tiny dielectric (non absorbing) and spherical particles, such as bubble, droplet, and air molecules (mainly nitrogen molecules, N_2) in the atmosphere between 10 and 20 km altitude whose dimensions are smaller than their wavelength, i.e., $\beta \ll 1$ [26]. The degree of scattering varies as a function of the ratio of the particle diameter to the wavelength of the radiation, along with other factors including polarization, angle, and coherence. The scattering coefficient is proportional to λ^{-1}, known as the Rayleigh law, which means that light of longer wavelength would experience less scattering as compared to light with shorter wavelength.

© 2015 by Taylor & Francis Group, LLC

(b) Mie scattering: Named after Gustav Mie, Mie scattering arises from dust, predominantly for particles larger than the Rayleigh range, i.e., $\beta \approx 1$. Scattering losses decrease rapidly with increasing wavelength. Mie scattering has a spectral dependence of approximately λ^{-1}. It is not strongly wavelength dependent, but produces the white glare around the Sun in the presence of particulate material in the air. It produces a pattern-like lobe, with a sharper and intense forward lobe for larger particles.

If the ratio of particle diameter to wavelength is more than $\beta \gg 1$, the laws of geometric optics are applied to describe the interaction of light with the particle.

2. Inelastic scattering: Inelastic scattering involves some changes in the energy of the radiation. The following categories of scattering are considered to be inelastic:

   (a) Brillouin scattering: Scattering of light from acoustic modes is called the Brillouin scattering[27], which is named after Léon Brillouin (1889−1969); the shift in frequency is $\Delta\bar{\nu} \leq 0.1\ cm^{-1}$. From a classical point, the compression of the medium changes the index of refraction and, therefore, leads to some reflection or scattering at any point where the index changes. From a quantum point of view, the process is considered one of interaction of light photons with acoustic or vibrational quanta (phonons). To note, a phonon represents an excited state in the quantum mechanical quantization of the modes of vibrations of elastic structures of interacting particles. When acoustic standing waves are produced in a solid, they create a periodic condition which may scatter light waves according to the Bragg law. This law states that when X-rays are scattered from a crystal lattice, peaks of scattered intensity are observed corresponding to the conditions, such as

      - the angle of incidence is equal to the angle of scattering and
      - the pathlength difference is equal to an integer number of wavelengths.

   (b) Compton scattering: A. H. Compton (1892−1962) observed the scattering of X-rays from electrons by solid materials (mainly graphite) and found the shift of the wavelength of the scattered photon. The incoming photons interact with the electron of the absorbing material and part of the photon energy is transferred to that electron, which gets scattered. This phenomenon is called the Compton scattering[28].

   (c) Raman scattering: Scattering in which the scattered photons have either a higher or lower photon energy is called the Raman scattering. In this case, the shift in frequency is $\Delta\bar{\nu} \leq 10^4\ cm^{-1}$. The

© 2015 by Taylor & Francis Group, LLC

incident photons interact with the molecules in a fashion that energy is gained or lost so that the scattered photons are shifted in frequency. Like Rayleigh scattering, Raman scattering depends on polarizability of the molecules.

1.2 What Is a Wave?

A wave may be described as a periodic disturbance that transports energy from one point to another. Its direction of propagation is the direction in which energy is carried. It is thought to be the result of correlated oscillations occurring at every point along the path of the wave. The direction of these oscillations determines the nature of a particular wave, for example,

1. Transverse wave: The oscillations are known as transverse when the vibration takes place in a direction perpendicular to the direction of propagation.
2. Longitudinal wave: In this case, the vibrations are in the direction of propagation.

For a sound wave, in which the pattern of disturbance caused by the movement of energy traveling through a medium as it propagates away from the source, the particles move back and forth parallel to the direction of the propagation of the wave.

The wave phenomena and their oscillatory motion possess many degrees of freedom. They are the most important physical feature of a given system. Systems that are much larger than the atomic scales are continuous at scales of small distances. The discrete atomic properties tend to be replaced by their continuous local averages. Their degrees of freedom are denoted by a function of space and time. The infinite continuous set of degrees of freedom is subjected to a set of mutually dependent nonlinear equations of motion. These equations are fused into a partial differential equation, whose temporal dependence and derivatives are inherited from the Newtonian "single atom" equation of motion or its relativistic generalization. The spatial dependence and derivatives of this equation are byproducts of the discrete atomic indices in the continuum limit.

A plane wave may be a simple two-dimensional (2D) or three-dimensional (3D) wave. Its characteristic is that all the points on a plane that is perpendicular to the direction of propagation have the same phase value. Waves propagating in 2D or 3D are analyzed using the concept of wavefront similar to ripples created by a stone when dropped in a pool of water. The wavefront is the locus of points of constant phase and the successive wavefronts are separated by one wavelength. For a 2D wave, the wavefronts are parallel lines,

© 2015 by Taylor & Francis Group, LLC

separated by a distance equal to the wavelength of the wave, while for a 3D wave, the wavefronts are parallel planes. The wavefronts connect points which have the same phase. They are perpendicular to the direction of the wave at every point. The wavefronts from a distant point source, say, a star, are circular, because the emitted light travels at the same speed in all directions. After traveling a long time, the radius of the wavefront becomes so large that the wavefronts are planar over the aperture of a telescope. The incident star beam is a stream of photons arriving at random times from a range of random angles within its angular diameter. The photon senses the presence of all the details of the collecting aperture. However, it is prudent to think of a wave, instead of a photon, as a series of wavelets (little waves) propagating outwards. The incident idealized photon is monochromatic in nature. The corresponding classical wave has the same extent as well.

For a wave traveling through a medium, a crest is seen moving along from particle to particle. This crest is followed by a trough which, in turn, is followed by the next crest. A distinct wave pattern in the form of a sine wave is observed traveling through the medium. This sine wave pattern continues to move in uninterrupted fashion until it encounters another wave along the medium or until it encounters a boundary with another medium. This type of wave pattern is referred to as a traveling wave; for instance, an ocean wave is falling under such category. The wave properties that are described by the following quantities are interrelated.

1. Amplitude: The amplitude of a wave is the maximum displacement of a particle from its equilibrium position as the wave passes through it (see Figure 1.3). It is measured in meters (m).

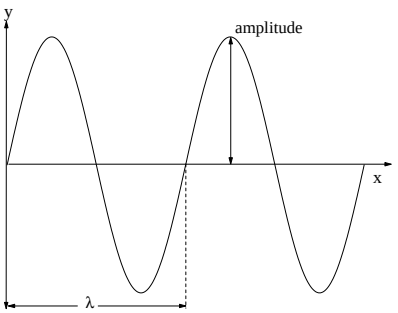


FIGURE 1.3: Amplitude pattern.

2. Frequency: The number of cycles per unit of time is called the frequency, ν, of oscillations caused by the wave. The unit of frequency is hertz (Hz; cycles per second). The quantity

$$\nu = \frac{\omega}{2\pi} = \frac{1}{T} \tag{1.1}$$

© 2015 by Taylor & Francis Group, LLC

where ω is the angular frequency, which is 2π times the frequency, ν, and T the period of the vibrations; one complete cycle of the wave is associated with an angular displacement of 2π radians.

The angular frequency, ω, of a wave is the number of radians per unit of time at a fixed position.

3. Path difference: The path length, l, is the distance through which a wavefront recedes when the phase increases by δ and is expressed as

$$l = \frac{v}{\omega}\delta = \frac{\lambda}{2\pi}\delta = \frac{\lambda_0}{2\pi n}\delta \tag{1.2}$$

where v is the velocity, λ the wavelength, λ_0 the wavelength in free space (vacuum),

$$n = \frac{c}{v} \tag{1.3}$$

the refractive index for refraction from vacuum into that medium, and c the speed of light in free space.

Let us denote n_1 and n_2 as the respective absolute refractive indices of first and second media. In such a situation, the relative refractive index, n_{12} for refraction from the first medium, into the second medium, would be

$$n_{12} = \frac{n_2}{n_1} = \frac{v_1}{v_2} \tag{1.4}$$

The value of v is not determined directly, but is relative to the speed of light in vacuum, c, with the help of the law of refraction. According to this law, if a plane electromagnetic wave falls onto a plane boundary between two homogeneous media, the sine of the angle, θ_1, bears a constant ratio to the sine of the angle, θ_2. Here θ_1 is the angle between the normal to the incident wave and the normal to the surface, and θ_2 the angle between the normal to the refractive wave and the surface normal, this constant ratio being equal to the ratio of the velocities, v_1 and v_2, of propagation in the two media:

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2} \tag{1.5}$$

Let us assume that the wavefront is continuous (though it has a kink at the boundary), so that the line of intersection between the incident wave and the boundary travels at the same speed, for instance, v', as the line of intersection between the refracted wave and the boundary. Therefore, we obtain

$$\begin{aligned} v_1 &= v' \sin\theta_1 \\ v_2 &= v' \sin\theta_2 \end{aligned} \tag{1.6}$$

© 2015 by Taylor & Francis Group, LLC

If we measure the distance traveled by two waves and compare those distances, any difference in the distances traveled is called the path difference. The path difference is measured in meters.

4. Period: Denoted by T, the period is the shortest interval in time between two instants when parts of the wave profile that are oscillating in phase pass a fixed point. If the periodic motion occurs ν times per second, the time for one cycle is $1/\nu$; hence,

$$T = \frac{1}{\nu} \tag{1.7}$$

A periodic traveling wave has a definite wavelength and a definite frequency, which are related to the velocity of the wave.

5. Phase: Any portion of a wave cycle is called a phase (optical path length), which is measured as an angle in degrees. When two waves of equal wavelength travel together in the same direction, they are said to be in phase if they are perfectly aligned in their cycle, and out of phase if they are out of step. In a function, $f(t) = a\cos(\omega t + \delta)$, in which the time, t, is a variable parameter, while the other quantities, namely, amplitude, $a(> 0)$; angular frequency, ω; and the phase shift, δ, are, in general, fixed and each of them influences the shape of the graph of this function. The phase shift represents a shift from zero phase, which occurs in the phase of one quantity or in the phase difference between two or more quantities. Here, the argument $\omega t + \delta$ of the cosine term is known as the phase. Unlike radio waves, the phase of optical waves cannot be measured directly due to the quantum nature of light dealt by quantum optics.

6. Velocity: The velocity, v, is defined to be the distance traveled of one wavelength by a wave in unit time period, T, that is,

$$v = \frac{\lambda}{T} = \nu\lambda = \frac{\omega}{|\,\boldsymbol{\kappa}\,|} \tag{1.8}$$

where κ denotes the wavenumber.

The velocity does not depend on the frequency of the disturbance. In a nondispersive medium, a pulse can travel down the medium without changing its shape. The frequency, ν, multiplied by the wavelength, λ, is equal to the velocity, which is the same for all frequencies. If the frequency increases, the wavelength decreases with the product being constant.

7. Wavelength: The least distance between two points on the wave profile that are oscillating in phase at a given instant of time is called the

© 2015 by Taylor & Francis Group, LLC

wavelength of the wave and is denoted by the symbol λ (see Figure 1.3), that is,

$$\lambda = v\frac{2\pi}{\omega} = vT \tag{1.9}$$

The units used for wavelength are in angstroms, $A = 10^{-10}$ meter (m), or nm $= 10^{-9}$ m, or $\mu m = 10^{-6}$ m. The wavelength that propagates in free space corresponding to harmonic wave of the same frequency is represented by λ_0 and is given by

$$\lambda_0 = cT = n\lambda \tag{1.10}$$

8. Wavenumber: The wavenumber is defined as the number of wavelength in vacuum per unit length (cm). This quantity is widely used in spectroscopy and is expressed as

$$\kappa = \frac{1}{\lambda_0} = \frac{\nu}{c} \tag{1.11}$$

The wavenumber, κ, is the number of radians per unit distance at a fixed time. It is convenient to define vectors $\boldsymbol{\kappa}_0$ and $\boldsymbol{\kappa}$ in the direction $\hat{\mathbf{s}}$ of propagation, whose respective lengths are

$$\kappa_0 = 2\pi\kappa = \frac{\omega}{c} = \frac{2\pi}{\lambda_0} \tag{1.12}$$

and

$$\kappa = n\kappa_0 = \frac{2\pi}{\lambda} = \frac{n\omega}{c} = \frac{\omega}{v} \tag{1.13}$$

9. Wave vector: The wavenumber vector $\boldsymbol{\kappa}$ is given by the relation

$$\boldsymbol{\kappa} = \frac{\omega}{c}\hat{\mathbf{s}} = \frac{2\pi}{\lambda}\hat{\mathbf{s}}, \tag{1.14}$$

in which $\hat{\mathbf{s}}(= s_x, s_y, s_z)$ is the unit vector in the direction of propagation of the wave, and the wave vector in vacuum is given by $\boldsymbol{\kappa}_0 = \kappa_0\hat{\mathbf{s}}$; the wave vector in a medium in the direction of propagation is expressed as $\boldsymbol{\kappa} = \kappa\hat{\mathbf{s}}$.

In the most general case, ω is dependent on the direction of propagation, $\boldsymbol{\kappa}$, as well as on the magnitude, $\kappa = |\boldsymbol{\kappa}|$. Here,

$$\boldsymbol{\kappa} = \kappa_x\mathbf{i} + \kappa_y\mathbf{j} + \kappa_z\mathbf{k} \tag{1.15}$$

© 2015 by Taylor & Francis Group, LLC

1.3 Harmonic Wave

The harmonic wave equations represent any type of wave disturbance that varies in a sinusoidal manner. This includes wave on a string, water waves that involve a combination of both longitudinal and transverse motions, and sound waves.

Light is an electromagnetic radiation, propagating disturbance involving space and time variation. A light source is characterized by its emission spectrum, degree of coherence, and radiant intensity. Due to the physical nature of the light generation process, the light sources can be divided into

1. Thermal source: The thermal sources transform part of their thermal energy into light. They have a continuous spectrum (see Section 1.1.2) and may generally be modeled with black body theory[29]. Sunlight or tungsten light sources are examples of such sources. A black body refers to an object or system that absorbs all radiation incident upon it and reradiates energy which is characteristic of this radiating system. The spectrum of electromagnetic energy emitted by a black surface is universal, which depends on the temperature of the surface and is independent of the size, shape or chemical composition of the material. For such sources, the spectrum depends only on the temperature (T). After knowing the temperature, one may calculate the optical power density and the spectral power density.

2. Luminescent light source: In luminescent sources, atoms emit and absorb photons at characteristic energies producing emission lines in the spectrum of each atom, which can be spontaneous such as light-emitting diodes (LED); emission can also be stimulated, viz., laser.

The harmonic variations of the electric and magnetic fields are always perpendicular to each other and to the direction of propagation denoted by $\boldsymbol{\kappa}$ (see Figure 1.4). The other notable features are

- the cross product $\mathbf{E}\times\mathbf{B}$ provides the direction of travel and
- the field always vary sinusoidally, with the same frequency and in phase with each other.

These variations in vacuum are described by the harmonic wave equations.

1.3.1 Harmonic Plane Waves

A harmonic plane wave represents a wave field spread out periodically in space and time. The propagation of light, $V(\mathbf{r}, t)$, stands for either of the varying field vectors that together constitute the wave. The quantity $V(\mathbf{r}, t)$ may refer to vertical displacements of a string or pressure variations due to

© 2015 by Taylor & Francis Group, LLC

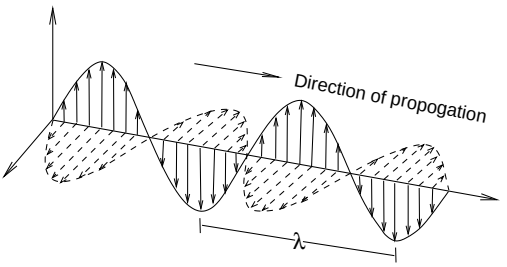


FIGURE 1.4: Plane electromagnetic wave.

a sound wave propagating in a gas. Let us examine the propagation of plane electromagnetic waves in a homogeneous medium in a region free of currents and charges. Each rectangular component $V(\mathbf{r},t)$ of the field vectors obeys the homogeneous wave equation

$$\left(\boldsymbol{\nabla}^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) V(\mathbf{r},t) = 0 \tag{1.16}$$

at each source-free point, in which V is an arbitrary function of its arguments, $V(\mathbf{r},t)$ the wave function describing the system (displacement), t the time, v the velocity of propagation of the oscillation and dependent on the physical features of the system, $\boldsymbol{\nabla}^2 (= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})$ the Laplacian operator (see equation B.14; Appendix B.1.4) in Cartesian coordinates, $j = x, y, z$, and

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \tag{1.17}$$

the position vector of a point, (x, y, z), in space.

The wave equation (1.16) is a linear second-order differential equation of a hyperbolic signature. In general, the differential equation is nonlinear, but in this case with the initial apptroximation it is assumed to be linear. The general solution of this equation in free space, in the form of $V = V(\mathbf{r}\cdot\hat{\mathbf{s}}, t)$, represents a plane wave in the direction given by the unit vector, $\hat{\mathbf{s}} (= s_x, s_y, s_z)$, since at each instant of time, V is constant over each of the planes, $\hat{\mathbf{s}}\cdot\mathbf{r} =$ constant, which are perpendicular to $\hat{\mathbf{s}}$. Let us choose a set of Cartesian axes $\mathsf{O}\xi$, $\mathsf{O}\eta$, $\mathsf{O}\zeta$ with $\mathsf{O}\zeta$ in the direction of $\hat{\mathbf{s}}$ (see Figure 1.5), so $\hat{\mathbf{s}}\cdot\mathbf{r} = s_x x + s_y y + s_z z = \zeta$ and we obtain

$$\frac{\partial}{\partial x} = s_x \frac{\partial}{\partial \zeta}; \qquad \frac{\partial}{\partial y} = s_y \frac{\partial}{\partial \zeta}; \qquad \frac{\partial}{\partial z} = s_z \frac{\partial}{\partial \zeta}$$

From these relations, we may write

$$\boldsymbol{\nabla}^2 V = \frac{\partial^2 V}{\partial \zeta^2} \tag{1.18}$$

© 2015 by Taylor & Francis Group, LLC

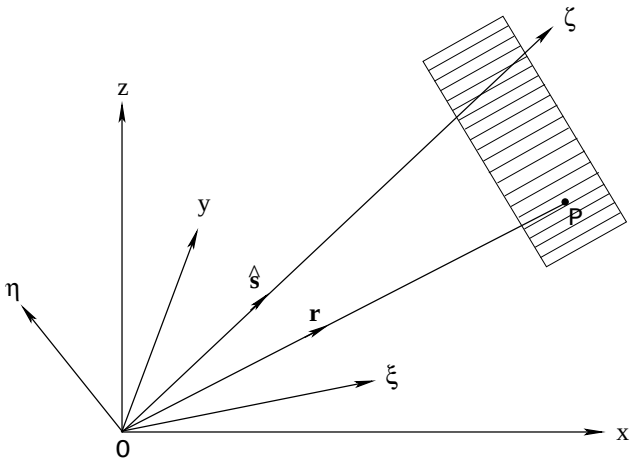


FIGURE 1.5: Propagation of a plane wave.

By simplifying the consideration to one-dimensional (1D) scalar waves in a vacuum, equation (1.16) takes the form

$$\left(\frac{\partial^2}{\partial\zeta^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) V = 0 \tag{1.19}$$

Let us set $\zeta - vt = p$ and $\zeta + vt = q$; therefore, equation (1.19) translates into

$$\frac{\partial^2 V}{\partial p \partial q} = 0 \tag{1.20}$$

In uni-dimensional space, the general solution of equation (1.20) in free, boundaryless space reduces to

$$\begin{aligned} V &= V_1(p) + V_2(q) \\ &= V_1(\hat{\mathbf{s}}\cdot\mathbf{r} - vt) + V_2(\hat{\mathbf{s}}\cdot\mathbf{r} + vt) \end{aligned} \tag{1.21}$$

where $V_1(\hat{\mathbf{s}}\cdot\mathbf{r} - vt)$ is a disturbance (or wave) traveling to the right propagating along the positive direction with velocity v, and similarly, $V_2(\hat{\mathbf{s}}\cdot\mathbf{r} + vt)$ is a wave traveling to the left.

1.3.2 Harmonic Spherical Waves

In a homogeneous medium, harmonic wave disturbances emanating from a point source spread out in all directions. These waves are considered to be spherical waves. The wavefronts corresponding to such waves are concentric surfaces and are perpendicular to the direction of propagation at any point. In contrast to the plane wave, where the amplitude remains constant as it propagates away from the source, the spherical wave decreases in amplitude.

© 2015 by Taylor & Francis Group, LLC

Let us reiterate that the strength of a wave is described by the intensity of the wave (see Section 1.5). For an isotropic source radiating its energy equally in all directions, the energy emitted per second by the source S passes through the surface of a sphere of area $4\pi r^2$ (see Figure 1.6), in which r is the radius of the sphere. If the energy is not absorbed, the energy flowing through the sphere per second is $I \times 4\pi r^2$. For a spherical wave, the intensity I is related to the distance from the source by an inverse square law. The energy per second reaching a detector transforms the light intensity into electrical signal, of fixed area decreasing as $1/r^2$, in which r is the distance of the detector from the source.

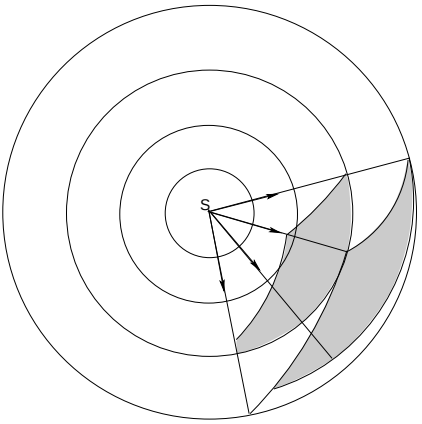


FIGURE 1.6: Propagation of a spherical wave.

Let us make the assumption that the function $V(\mathbf{r}, t)$ has spherical symmetry about the origin, i.e., $V(\mathbf{r}, t) = V(r, t)$, in which $r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$. Using the relations

$$\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} = \frac{x}{r}\frac{\partial x}{\partial r}, \text{ etc.} \tag{1.22}$$

after a calculation we get

$$\boldsymbol{\nabla}^2 V = \frac{1}{r}\frac{\partial^2}{\partial r^2}[rV] \tag{1.23}$$

thus, the wave equation (1.16) has the form

$$\left(\frac{\partial^2}{\partial r^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) rV = 0 \tag{1.24}$$

The general solution may be written as

$$V = \frac{1}{\mathbf{r}}V_1(r - vt) + \frac{1}{\mathbf{r}}V_2(r + vt) \tag{1.25}$$

© 2015 by Taylor & Francis Group, LLC

The first term on the right-hand side (RHS) of equation (1.25) represents a spherical wave which diverges from the origin, while the second term converges towards the origin.

1.3.3 Fourier Transform Method

A field with general time dependent is a linear superposition of fields that vary harmonically with time at different frequencies. This relationship is known as the Fourier transform (FT), named after J. B. J. Fourier (1768−1830), which is used extensively to analyze the output from optical as well as radio telescopes. It is based on the discovery that it is possible to take any periodic function of time $f(t)$ and transform it to a function of frequency $\widehat{f}(\omega)$, in which the notation, $\widehat{\ }$, denotes the FT of a particular physical quantity; the frequency ω and the time t are Fourier dual coordinates (see Figure 1.7). The linear addition of various components of a quantity is carried out by summing up the phasor vectors. Adding two components with identical phasors but with opposite signs of the angular frequency, ω, provides the real quantity. The free wave equation is a linear homogeneous differential equation; therefore, any linear combination of its solution is a solution as well. The Fourier transform pair with the harmonic function of frequency[106] in time domain, t, can be expressed as

$$\widehat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt \tag{1.26}$$

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}(\omega)e^{i\omega t}d\omega \tag{1.27}$$

Let us consider the 1D wave equation

$$\left(\frac{\partial^2}{\partial\zeta^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)V(\zeta,t) = 0 \tag{1.28}$$

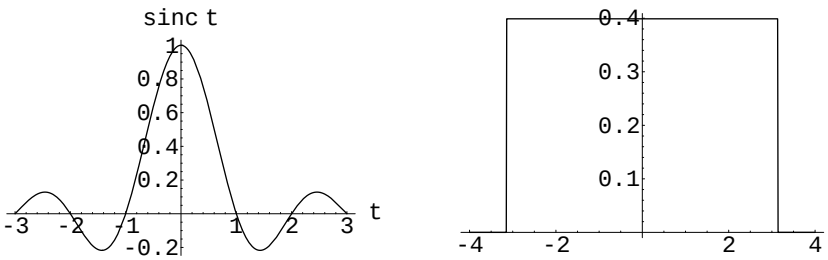


FIGURE 1.7: Fourier transform of $sinc\, t$.

Following equations (1.26 and 1.27), the Fourier transform pair for $V(\zeta,t)$

© 2015 by Taylor & Francis Group, LLC

are given by

$$V(\zeta,\omega) = \int_{-\infty}^{\infty} V(\zeta,t)e^{-i\omega t}dt \tag{1.29}$$

$$V(\zeta,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{V}(\zeta,\omega)e^{i\omega t}d\omega \tag{1.30}$$

Therefore, we write

$$\frac{\partial^2 V(\zeta,t)}{\partial\zeta^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\partial^2\widehat{V}(\zeta,\omega)e^{i\omega t}}{\partial\zeta^2}d\omega \tag{1.31}$$

$$\frac{\partial^2 V(\zeta,t)}{\partial t^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{V}(\zeta,\omega)(-\omega^2)e^{i\omega t}d\omega \tag{1.32}$$

so equation (1.28) transforms into

$$\left(\frac{\partial^2}{\partial\zeta^2} + \frac{\omega^2}{v^2}\frac{\partial^2}{\partial t^2}\right)\widehat{V}(\zeta,\omega) = 0 \tag{1.33}$$

Equation (1.33) is recognized as the equation of a harmonic oscillator whose solution is

$$\widehat{V}(\zeta,\omega) = A(\omega)e^{i\kappa\zeta} + B(\omega)e^{-i\kappa\zeta} \tag{1.34}$$

in which $\kappa(=\omega/v)$ is the wavenumber and $A(\omega)$ and $B(\omega)$ are the constants of integration.

The solution for equation (1.28) can be found by substituting $\widehat{V}(\zeta,\omega)$ in (1.34) into FT $V(\zeta,t)$ in equation (1.30)

$$\begin{aligned} V(\zeta,t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[A(\omega)e^{i\kappa\zeta} + B(\omega)e^{-i\kappa\zeta}\right]e^{i\omega t}d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} A(\omega)e^{i\omega(t+\zeta/v)}d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} B(\omega)e^{i\omega(t-\zeta/v)}d\omega \end{aligned} \tag{1.35}$$

From equations (1.29 and 1.30), we get

$$V(\zeta,t) = V_1\left(t+\frac{\zeta}{v}\right) + V_2\left(t-\frac{\zeta}{v}\right) \tag{1.36}$$

which is equivalent to the solution given in equation (1.21). The Fourier transform pair provide us the solution of the equation as

$$V(\zeta,t) = A\sin(\omega t+\kappa\zeta) + B\sin(\omega t-\kappa\zeta) \tag{1.37}$$

There are two more ways of introducing time harmonic fields, for example,

© 2015 by Taylor & Francis Group, LLC

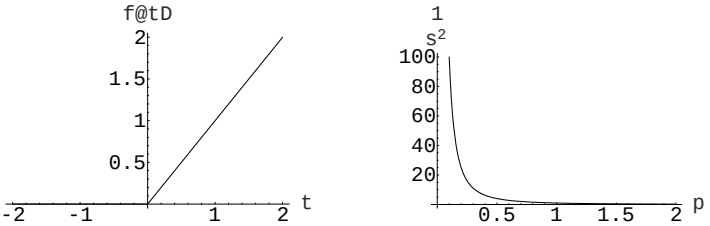


FIGURE 1.8: Laplace transform of Heaviside unit step.

1. Laplace transform: The Laplace (P. S de Laplace; 1749−1827) transform is an integral transform and is useful in solving linear ordinary differential equations. It simplifies the system analysis and maps a function in the time domain, $f(t)$, defined on $0 \leq t < \infty$ to a complex function (see Figure 1.8)

$$F(s) = \mathcal{L}\{f(t)\} = \int_0^\infty f(t)e^{-st}dt \tag{1.38}$$

$$f(t) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} F(s)e^{st}ds \tag{1.39}$$

in which $\mathcal{L}$ stands for the Laplace transform operator and s the complex quantity, which demands a suitable contour of integration to be defined on the complex s plane.

2. Real Value Convention: This is purely harmonic, $\cos\omega t$, which preserves units, for instance,

$$f(t) = \Re\left\{f(\omega)e^{i0\omega t}\right\} \tag{1.40}$$

1.3.4 Phase Velocity

The phase velocity of a wave is the rate at which the phase of the wave propagates in space. This is the speed at which the phase of any one frequency component of the wave travels. For such a component, any given phase of the wave, for instance, the crest, would appear to travel at the phase velocity. In the simple case of a pure traveling sinusoidal wave, let us imagine a rigid profile being physically moved in the positive x-direction with speed v. As stated in Section 1.2, the wave function depends on both time and position. At any fixed instant of time, the function varies sinusoidally along the x axis, whereas at any fixed location on the x axis, the function varies sinusoidally with time. We may express a traveling wave as the function

$$V(z,t) = A\cos(\kappa z - \omega t) \tag{1.41}$$

where A is the amplitude.

© 2015 by Taylor & Francis Group, LLC

The function $V(z,t)$ is the solution of the 1D wave equation

$$\frac{\partial^2 V}{\partial z^2} - \left(\frac{\kappa}{\omega}\right)^2 \frac{\partial^2 V}{\partial t^2} = 0 \tag{1.42}$$

Since ω is the number of radians of the wave that pass a given location per unit of time and $1/\kappa$ is the spatial length of the wave per radian, it follows that

$$v_p = \frac{\omega}{\kappa} \tag{1.43}$$

where v_p is the phase velocity.

The phase velocity is the speed at which the shape of the wave is moving, i.e., the speed at which any fixed phase of the cycle is displaced.

1.3.5 Group Velocity

The group velocity of the wave is defined as the velocity at which the overall shape of the waves' amplitudes (group of waves), called the modulation (or envelope) of the wave, propagates through space. It follows from the Fourier theorem (see Section 1.3.3) that any wave, $V(\mathbf{r},t)$, may be regarded as a superposition of monochromatic waves (see Section 1.4.1) of different frequencies. A wave where a single wavelength is present is called the monochromatic (see Section 1.4), unlike white light in which are present all wavelengths in the visual range. It can be considered to be near-monochromatic, if the Fourier amplitudes, $a(\omega)$, differ from zero within a narrow range

$$\bar{\omega} - \frac{1}{2}\Delta\omega \le \omega \le \bar{\omega} + \frac{1}{2}\Delta\omega \qquad \frac{\Delta\omega}{\bar{\omega}} \ll 1 \tag{1.44}$$

Let us consider that a wave is formed by the superposition of two plane monochromatic waves of equal amplitude, but neighboring frequencies and wave numbers are propagating in the direction of the z axis. Differences in frequency imply differences in wavelength and velocity. The superposition of these two waves, with wave crests moving at different speeds, exhibits periodically large and small amplitudes. The resultant wave is expressed as

$$\begin{aligned} U(z,t) &= Ae^{i(\kappa_1 z - \omega_1 t)} + Ae^{i(\kappa_2 z - \omega_2 t)} \\ &= A\left(e^{i\{[(\kappa_1-\kappa_2)/2]z - [(\omega_1-\omega_2)/2]t\}} + e^{i\{[(\kappa_2-\kappa_1)/2]z - [(\omega_2-\omega_1)/2]t\}}\right) \\ &\quad \times e^{i[(\kappa_2-\kappa_1)/2]z - i[(\omega_2-\omega_1)/2]t} \\ &= 2A\cos\left(\frac{\kappa_1-\kappa_2}{2}z - \frac{\omega_1-\omega_2}{2}t\right) e^{i\{[(\kappa_1+\kappa_2)/2]z - [(\omega_1+\omega_2)/2]t\}} \\ &= 2A\cos(\kappa_g z - \omega_g t)e^{i(\bar{\kappa}z - \bar{\omega}t)} \end{aligned} \tag{1.45}$$

with

$$\begin{aligned} \bar{\kappa} &= \frac{\kappa_1+\kappa_2}{2}, \qquad \bar{\omega} = \frac{\omega_1+\omega_2}{2} \\ \kappa_g &= \frac{\kappa_1-\kappa_2}{2}, \qquad \omega_g = \frac{\omega_1-\omega_2}{2} \end{aligned} \tag{1.46}$$

© 2015 by Taylor & Francis Group, LLC

as the mean frequency and the mean wave number, respectively.

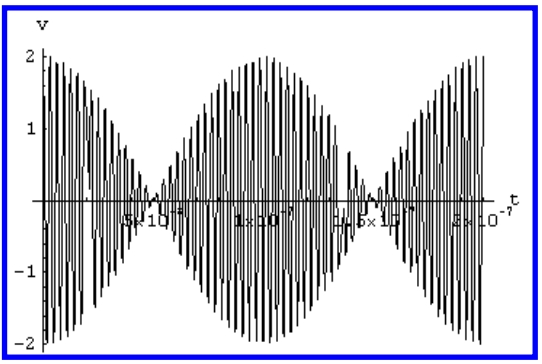


FIGURE 1.9: One-dimensional plot (in time domain) of the cosine factor of $V(z,t) = 2A\cos(\bar{\kappa}z - \bar{\omega}t)\cos(\kappa_g z - \omega_g t)$; the low frequency wave serves as an envelope modulating the high frequency wave. The dashed lines of this curve display the envelope of the resulting wave disturbance.

By this transformation the wave is split into an amplitude factor slowly oscillating at $\omega_1 - \omega_2$ and a phase factor rapidly oscillating at $\omega_1 + \omega_2$. Equation (1.45) is interpreted as representing a plane wave of frequency, $\bar{\omega}$, and wavelength, $2\pi/\bar{\kappa}$, propagating in the z-direction. The amplitude of this wave varies with time and position, between $2A$ and 0 (see Figure 1.9), exhibiting the phenomenon of beats. The beat is produced by the superposition of two equal amplitude harmonic waves of different frequency, which can be perceived as periodic variations in volume whose rate is the difference between the two frequencies. The successive maxima of the amplitude function are at intervals

$$\delta t = \frac{4\pi}{\delta\omega}\,(\text{with z fixed}) \quad \text{or} \quad \delta z = \frac{4\pi}{\delta\kappa}\,(\text{with t fixed}) \tag{1.47}$$

from each other, while the maxima of the phase function are at intervals

$$\delta t = \frac{2\pi}{\bar{\omega}}\,(\text{with z fixed}) \quad \text{or} \quad \delta z = \frac{2\pi}{\bar{\kappa}}\,(\text{with t fixed}) \tag{1.48}$$

Since $\delta\omega/\bar{\omega}$ and $\delta\kappa/\bar{\kappa}$ are assumed to be small compared with unity, the amplitude would vary slowly in comparison with the other term. It follows from equation (1.45) that the planes of constant amplitude and the maxima of the amplitude propagate with velocity, v_g ([142], [107])

$$v_g = \frac{\delta\omega}{\delta\kappa} \tag{1.49}$$

while the planes of constant phase propagate with the phase velocity, $v_p = \bar{\omega}/\bar{\kappa}$ (see equation 1.43). Since the differences between the frequencies and

© 2015 by Taylor & Francis Group, LLC

propagation constants are small, $\delta\omega/\delta\kappa$ may be replaced with $d\omega/d\kappa$, so that expression for the group velocity is written in the form

$$v_g = \frac{d\omega}{d\kappa} = \frac{\omega_g}{\kappa_g} \tag{1.50}$$

The propagation of information or energy in a wave always occurs as a change in the wave. We have differentiated the dispersion relation with respect to κ, which is valid under general conditions. The physical meaning of equation (1.50) is that the wave packet (a smooth envelope times an oscillatory carrier wave) with parameter (ω, κ) would move approximately at the velocity, v_g. The same holds for a wavefront and for any other signal that can carry information. The energy of the wave is determined by its amplitude, so in general, the group velocity provides the velocity of the energy transport. Let us plug equation (1.43) into equation (1.49), so that the relation between group and phase velocities is derived as

$$\begin{aligned} v_g &= \frac{d\omega}{d\kappa} = \frac{d}{d\kappa}\left(dv_p\right) \\ &= v_p + \kappa\left(\frac{dv_p}{d\kappa}\right) = v_p - \lambda\left(\frac{dv_p}{d\lambda}\right) \end{aligned} \tag{1.51}$$

In a dispersion-free medium, $dv_p/d\kappa = 0$, equation (1.51) turns out to be $v_g = v_p$. This is the case of light propagating in vacuum, where $v_g = v_p = c$.

1.3.6 Dispersion Relation

The dispersion relation elucidates the effect of dispersion in a medium on the properties of a wave traveling within that medium. Since V obeys the wave equation, the frequency and the wave number are related. By substituting equation (1.45) into wave equation (1.16), the following relationship emerges

$$\begin{aligned} \left(\boldsymbol{\nabla}^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) Ae^{i(\omega t - \boldsymbol{\kappa}\cdot\mathbf{r})} &= -\left(\boldsymbol{\kappa}^2 - \frac{\omega^2}{v^2}\right) Ae^{i(\boldsymbol{\kappa}\cdot\mathbf{r} - \omega t)} \\ &= -\left(\boldsymbol{\kappa}^2 - \frac{\omega^2}{v^2}\right) U = 0 \end{aligned} \tag{1.52}$$

where $\boldsymbol{\nabla}^2$ is the Laplacian operator in Cartesian coordinates (see equation B.14; Appendix B.1.4).

Thus,

$$\boldsymbol{\kappa}^2 = \frac{\omega^2}{v^2} \tag{1.53}$$

The wave velocity in equation (1.53) provides the dispersion relation. In general, this is a functional relation between the frequency, ω, and the wave vector, $\boldsymbol{\kappa}$, and is given by $\omega = \omega(\boldsymbol{\kappa})$. If the dispersion relation is linear, equations (1.43) and (1.49) become equal and the system is referred to as nondispersive.

© 2015 by Taylor & Francis Group, LLC

In a transparent optical medium, the refractive index, n, is defined as the ratio c/v_p, in which c is the speed of light in free space and v_p the phase velocity of light in that medium. Since $v_p = \omega/\kappa$, we have $\omega = \kappa c/n$. If the dispersion relation is nonlinear, the system is dispersive. In the case of light waves through dielectric media, the dependence of the refractive index (see Section 1.2), $n(\omega)$, on the wave frequency causes dispersion, i.e., $\kappa = n(\omega)(\omega/c)$; hence, the phase velocity is expressed as

$$v_p = \frac{c}{n(\omega)} = \frac{\omega}{\kappa} \tag{1.54}$$

and thus,

$$\frac{dv_p}{d\kappa} = \frac{d}{d\kappa}\left(\frac{c}{n(\omega)}\right) = \frac{-c}{n^2(\omega)}\left(\frac{dn(\omega)}{d\kappa}\right)$$

Invoking equation (1.51), we get

$$\begin{aligned} v_g &= v_p\left[1 - \frac{\kappa}{n(\omega)}\left(\frac{dn(\omega)}{d\kappa}\right)\right] \\ &= v_p\left[1 + \frac{\lambda}{n(\omega)}\left(\frac{dn}{d\lambda}\right)\right] \\ &= \frac{v_p}{1 - \dfrac{\lambda}{n(\omega)}\dfrac{dn(\omega)}{d\lambda}} \qquad \text{if } \frac{\lambda}{n(\omega)}\frac{dn(\omega)}{d\lambda} \ll 1 \end{aligned} \tag{1.55}$$

where $\kappa = 2\pi/\lambda$ and $d\kappa/d\lambda = -\left(2\pi/\lambda^2\right)$.

Hence,

$$\frac{c}{v_g} = \frac{c}{v_p}\left[1 - \frac{\lambda}{n(\omega)}\frac{dn(\omega)}{d\lambda}\right]$$

that is,

$$\frac{c}{v_g} = n(\omega) - \lambda\frac{dn(\omega)}{d\lambda} \tag{1.56}$$

Equation (1.56) shows the relationship between dispersion and group velocity[173]. It is seen from this expression, and also from equation (1.51), that if $dn(\omega)/d\kappa = 0$, the group velocity turns out to be equal to the phase velocity, i.e., $v_g = v_p$. The dispersion of a medium for any kind of wave is, in general, negative, i.e., $dn(\omega)/d\kappa < 0$, except in short regions where it is positive. However, if $dn(\omega)/d\kappa$ is positive, this is known as a region of anomalous dispersion.

In the amplitude modulation (AM) of radio waves, the carrier waves are modulated to contain information. Here the group velocity, known as signal velocity, is normally less than the phase velocity of the carrier waves. If the light pulses are transmitted through a dispersive medium, the group velocity is the velocity of the pulses and will be different from the velocity of the individual harmonic waves. In the case of a nontrivial dispersion, for example,

© 2015 by Taylor & Francis Group, LLC

the propagation of electromagnetic radiation through an ionized plasma, the relation is provided by

$$\omega^2 = c^2\kappa^2 + \bar{\omega}^2 \tag{1.57}$$

This equation describes the relation between the energy (ω) and the momentum (κ) of massive relativistic particles in relativistic quantum mechanics. In equation (1.57), ω never drops below ω_p, which serves as the cutoff frequency. To note, a plasma is typically an ionized gas, the fourth state of matter, and is considered to be a distinct phase of matter in contrast to solids, liquids, and gases because of its unique characteristic that deals with electric charges. It consists of a collection of free moving electrons and ions. The free electric charges make the plasma electrically conductive. The term *ionized* means that at least one electron has been dissociated from an atom.

1.4 Monochromatic Fields

Monochromatic radiation is the radiation of single precise energy; the energy is related with its wavelength which is often used to specify the color of the visible radiation. Completely monochromatic radiation cannot be produced, but a quantum optical instrument, called a laser (Light Amplification by Stimulated Emission of Radiation) can produce a nearly monochromatic radiation (single wavelength) and a coherent beam of light by exciting atoms to a higher energy level and causing them to radiate their energy in phase[30]. A laser makes use of mechanisms, such as absorption, stimulation, and spontaneous emission ([165], [162] and references therein). The laser output can be continuous or pulsed and contributes significantly to the science, communications, and technology. Solid state lasers emit the ultrafast pulses, whose time durations are of the order of picoseconds (ps; 10^{-12} s) or femtoseconds (fs; 10^{-15} s). These lasers are used to map the sequence of the events, while tunable lasers are useful in strong-field interactions.

1.4.1 Complex Representation

Thus far we have noticed that the term $V(\mathbf{r}, t)$ is periodic both in $\boldsymbol{\kappa}\cdot\mathbf{r}$ and t, and it describes a wave propagating along the $\boldsymbol{\kappa}$ direction at a velocity of $v = \omega/\kappa$. A general time harmonic wave of frequency ω may be defined as a real solution of the wave equation, $V(\mathbf{r}, t)$, at a point, $\mathbf{r}$, of the form

$$V(\mathbf{r}, t) = \Re\left\{a(\mathbf{r}, \nu)e^{-i[2\pi\bar{\nu}t - \psi(\mathbf{r},\nu)]}\right\} \tag{1.58}$$

where $\bar{\nu}$ is the mean optical frequency, $\psi(\mathbf{r}, \nu)$ the phase function, $a(\mathbf{r}, \nu)$ the maximum amplitude of the field, which is complex-valued function of space.

© 2015 by Taylor & Francis Group, LLC

$\Re$ is the 'real part of', $\mathbf{r}(= x, y, z)$ the position vector of a point, and $\nu = 1/T$ represents the number of oscillations cycles in a unit time.

The phase component gains significance when we superpose several waves. The oscillations of V in equation (1.58) are bounded by $0 \leq |V| \leq a$ or $-a \leq V \leq a$. The most physically relevant information is embodied in the relative phase differences among superposed waves and these relative amplitude ratios, so that we may drop the $\Re$ symbol in equation (1.58). This information is encoded in the complex exponential representation. Denoting the complex amplitude of the vibration, $\mathbf{A}(\mathbf{r}, \nu)$, by

$$\mathbf{A}(\mathbf{r}, \nu) = a(\mathbf{r}, \nu) e^{i\psi(\mathbf{r}, \nu)} \tag{1.59}$$

therefore, the complex representation of the analytic signal of a plane wave, $U(\mathbf{r}, t)$, becomes

$$U(\mathbf{r}, t) = a(\mathbf{r}, \nu) e^{-i[2\pi\bar{\nu}t - \psi(\mathbf{r}, \nu)]} \tag{1.60}$$

$$= \mathbf{A}(\mathbf{r}, \nu) e^{-i2\pi\bar{\nu}t} \tag{1.61}$$

This complex representation is preferred for linear time invariant systems, because the eigenfunctions of such systems are of the form $e^{-i\bar{\omega}t}$, in which $\bar{\omega} = 2\pi\bar{\nu}$ is the angular frequency. The complex representation of the analytic signal of a spherical wave is represented by

$$U(\mathbf{r}, t) = \left[\frac{a(\mathbf{r}, \nu)}{\mathbf{r}}\right] e^{-i[2\pi\bar{\nu}t - \psi(\mathbf{r}, \nu)]} \tag{1.62}$$

Equation (1.62) is the solution to the wave equation and represents spherical wave propagating outward from the origin. The irradiance of the spherical wave is proportional to the square of the amplitude $a(\mathbf{r}, \nu)/\mathbf{r}$ at a distance $\mathbf{r}$.

The complex amplitude is a constant phasor in the monochromatic case; therefore, the Fourier transform (FT; see Section 1.3.3) of the complex representation of the signal, $U(\mathbf{r}, t)$, is given by

$$\widehat{U}(\mathbf{r}, \nu) = a(\mathbf{r}, \nu) e^{i\psi} \delta(\nu - \bar{\nu}) \tag{1.63}$$

in which $\widehat{\ }$ stands for a Fourier transform.

The spectrum is equal to twice the positive part of the instantaneous spectrum, $\widehat{V}(\mathbf{r}, \nu)$.

1.4.2 Superposition Principle

The superposition principle is the most important property of wave phenomena. The effects of interference and diffraction are the corollaries of this principle[153]. If two or more waves meet at a point in space, the net disturbance at the point at each instant of time is defined by the sum of the disturbances created by each of the waves individually, that is,

$$V = V_1 + V_2 + \cdots \tag{1.64}$$

© 2015 by Taylor & Francis Group, LLC

where $V_{j=1,2,\ldots}$ are the constituent waves.

Since the free wave equation is a linear homogeneous differential equation (1.16), any linear combination of its solution is a solution as well. The superposition of electromagnetic waves in terms of their electric field, **E**, and magnetic field, **B**, may also be expressed as

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \cdots \tag{1.65}$$

$$\mathbf{B} = \mathbf{B}_1 + \mathbf{B}_2 + \cdots \tag{1.66}$$

This type of linear superposition is valid in the presence of matter, however, deviations from linearity are observed at high intensities produced by lasers when the electric fields approach the electric fields comparable to atomic fields (nonlinear optics). Let the two waves, E_1, and E_2, of the same frequency, differing in amplitude and phase, combine to form a resultant wave, E; the orientation of the electric or magnetic fields must be taken into account. The time variations of these waves are expressed as

$$E_1(z,t) = a_1 \cos(\kappa z - \omega t + \delta_1) = \Re\left[a_1 e^{i\delta_1} e^{\kappa z - \omega t}\right] \tag{1.67}$$

$$E_2(z,t) = a_2 \cos(\kappa z - \omega t + \delta_2) = \Re\left[a_2 e^{i\delta_2} e^{\kappa z - \omega t}\right] \tag{1.68}$$

These two waves (equations 1.67 and 1.68) intersecting at a fixed point in space may differ in phase $\kappa\cdot(\mathbf{r}_2 - \mathbf{r}_1) + (\delta_2 - \delta_1)$. The difference in phase is due to a path difference provided by the first term, and an initial phase difference in the range $-\pi < \delta \leq \pi$ is given by the second term, $\delta = \delta_2 - \delta_1$. The real part is avoided for convenience while deriving the resultant wave,

$$\begin{aligned} E_0(z,t) &= E_1(z,t) + E_2(z,t) \\ &= a_1 e^{i\delta_1} e^{\kappa z - \omega t} + a_2 e^{i\delta_2} e^{\kappa z - \omega t} \\ &= \left(a_1 e^{i\delta_1} + a_2 e^{i\delta_2}\right)\left(e^{i(\kappa z - \omega t)}\right) \\ &= a_0 e^{i\delta_0} e^{i(\kappa z - \omega t)} \end{aligned} \tag{1.69}$$

Thus, the superposition of two harmonic waves of given frequency produces a harmonic wave of the same frequency with a given amplitude and phase. The amplitude and phase of the resultant wave are derived as

$$a_0 e^{i\delta_0} = a_1 e^{i\delta_1} + a_2 e^{i\delta_2} \tag{1.70}$$

or

$$a_0 \cos\delta_0 + i a_0 \sin\delta_0 = a_1 \cos\delta_1 + a_2 \cos\delta_2 + i\left(a_1 \sin\delta_1 + a_2 \sin\delta_2\right) \tag{1.71}$$

Here a_0 is the amplitude and $\delta_{j=1,2}$ the phase. The amplitude of the resultant wave is calculated as

$$\begin{aligned} a_0^2 &= Abs\left[a_0 e^{i\delta_0}\right]^2 = Abs\left[a_1 e^{i\delta_1} + a_2 e^{i\delta_2}\right]^2 \\ &= a_1^2 + a_2^2 + 2a_1 a_2 \cos(\delta_2 - \delta_1) \\ &= a_1^2 + a_2^2 + 2a_1 a_2 \cos\delta \end{aligned} \tag{1.72}$$

© 2015 by Taylor & Francis Group, LLC

Equation (1.72) depicts the expression of two-beam interference. If $a_1 = a_2$, we find

$$a_0^2 = 2a_1^2\,(1 + \cos\delta) = 4a_1^2 \cos^2\left(\frac{\delta}{2}\right) \tag{1.73}$$

Thus, the linear combination provides a cosine intensity distribution or equivalently a cosine squared distribution. Since the real and imaginary parts must be equal, the phase angle is determined by

$$\tan\delta_0 = \frac{a_0 \sin\delta_0}{a_0 \cos\delta_0} = \frac{a_1 \sin\delta_1 + a_2 \sin\delta_2}{a_1 \cos\delta_1 + a_2 \cos\delta_2} \tag{1.74}$$

For the combination of several sine waves, the resultant wave is given by

$$E(z,t) = \sum_{i=1}^{N} a_i e^{i\delta_i} e^{i(\kappa z - \omega t)} = a_0 e^{i\delta_0} e^{i(\kappa z - \omega t)} \tag{1.75}$$

or

$$a_0 e^{i\delta_0} = a_i e^{i\delta_i} \tag{1.76}$$

The phase of this resultant wave is derived as

$$\tan\delta_0 = \frac{\sum_{i=1}^{N} a_i \sin\delta_i}{\sum_{i=1}^{N} a_i \cos\delta_i} \tag{1.77}$$

The resultant amplitude is calculated as

$$\begin{aligned} a_0^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N} a_i e^{i\delta_i}\, a_j e^{i\delta_j} \\ &= \sum_{i=1}^{N} a_i^2 + \sum_{i=1}^{N}\sum_{i\neq j}^{N} a_i a_j e^{i(\delta_i - \delta_j)} \\ &= \sum_{i=1}^{N} a_i^2 + \sum_{i=1}^{N}\sum_{j>i}^{N} a_i a_j \left(e^{i(\delta_i - \delta_j)} + e^{-i(\delta_i - \delta_j)}\right) \\ &= \sum_{i=1}^{N} a_i^2 + 2\sum_{i=1}^{N}\sum_{j>i}^{N} a_i a_j \cos(\delta_i - \delta_j) \end{aligned} \tag{1.78}$$

Equation (1.78) expresses a harmonic wave of same frequency in which the last term is the interference term. If the number of randomly phased sources of equal amplitude is large, the waves are incoherent, and the sum of cosine terms approaches zero. The resultant irradiance of N identical but randomly phased sources is the sum of the individual irradiances, i.e.,

$$a_0^2 = \sum_{i=1}^{N} a_i^2 \tag{1.79}$$

© 2015 by Taylor & Francis Group, LLC

If all the a_i are equal, we get

$$a_0^2 = N a_0^2 \tag{1.80}$$

While for the in-phase coherent case, equation (1.78) turns out to be

$$a_0^2 = \sum_{i=1}^{N} \sum_{j=i}^{N} a_i a_j = \left[\sum_{i=1}^{N} a_i \right]^2 \tag{1.81}$$

Here all δ are assumed to be equal. If all the amplitudes are equal, we find

$$a_0^2 = N^2 a_0^2 \tag{1.82}$$

The resultant irradiance of N identical coherent sources, radiating in phase with each other, is equal to N^2 times the irradiance of the individual sources. For both coherent and incoherent cases, the total energy does not change, but the distribution of the energy does change.

1.4.3 Standing Wave

Standing waves (also called the stationary waves) are set up as a consequence of superposition of two waves of same amplitude and frequency propagating at the same speed in a unidimensional space, but in opposite directions. Unlike the traveling waves, standing waves transmit no energy. Standing waves are formed in a medium that has boundaries where a wave is reflected back. The waves bounce against these boundaries and partly reflect back into the medium and partly transmit outside the medium. The reflected waves superimpose on incident waves, which may, in turn, cancel each other. Having boundary means that the waves are confined to a specific length of the medium. The waves are moving, but the same places have a very large amplitude oscillation while others have zero amplitude and continuous destructive interference. The stationary waves may be set up when a wave reflects back from a surface and the reflected wave interferes with the wave still traveling in the original direction. The reflected wave and the incoming wave interfere. At the reflecting surface, the two waves are equal but opposite and get canceled out. Such a place is called a node of the wave. At other points along the waves, these waves the same. Therefore, they add together or interfere constructively. Such points are called the anti-nodes of the wave. If the boundary is closed, there will be a node at the boundary, while if the boundary is open, there will be an anti-node at the boundary. From the principle of superposition, the resultant wave, $U(z,t)$, is given by

$$\begin{aligned} U(z,t) &= a e^{i(\kappa z - \omega t + \delta_1)} + a e^{i(-\kappa z - \omega t + \delta_2)} && (1.83) \\ &= 2a e^{-i\omega t} \cos(\kappa z + \delta) && (1.84) \end{aligned}$$

For $\delta = \pi/2$, we have

$$V(z,t) = \Re[U(z,t)] = 2a \sin \kappa z \cos \omega t \tag{1.85}$$

© 2015 by Taylor & Francis Group, LLC

Equation (1.85) represents the standing wave and provides a solution of the unidirectional wave equation for waves which propagate along a bounded unidirectional region and interfere destructively at its end points finally vanishing. The first term of the RHS is space dependent, while the second term is time dependent. At any point x along the medium, the oscillations are given by $E = a\cos\omega t$, where $a = 2E_0 \sin\kappa z$. The wave function vanishes at the points

$$x_n = \frac{nx}{\kappa} = \frac{n\lambda}{2}, \qquad n = 0, \pm 1, \pm 2, \cdots \tag{1.86}$$

for any value of t. The places where $\sin(\kappa z)$ are zero would not move. These points are the nodes that are separated by half a wavelength. At a given time t, the function E obtains its extreme at the points

$$x_m = \left(m + \frac{1}{2}\right)\frac{\pi}{\kappa} = \left(m + \frac{1}{2}\right)\frac{\lambda}{2} \tag{1.87}$$

The places $\sin(\kappa z)$ are ± 1 would have the largest amplitudes. These positions are the anti-nodes. The distance between a node and an adjacent anti-node is one-fourth of a wavelength, i.e., $\lambda/4$. For a standing wave to exist in a region of length L, an appropriate number of quarter wavelengths, $\lambda/4$ should fit into the boundary, and the boundary conditions must be satisfied.

1.4.4 Doppler Effect

Doppler effect, named after Christian Andreas Doppler (1803−1853), is the apparent shift in frequency and wavelength of a light wave that is perceived by an observer moving with respect to the object. He used this concept to explain the color of binary stars[31].

1.4.4.1 Doppler Shift

The Doppler shift of the wavelength of light due to the velocity of the source is given by

$$\lambda_o = \lambda \frac{1 + v_r/c}{\sqrt{1 - v_s^2/c^2}} \tag{1.88}$$

in which λ_o is the observed wavelength, λ the wavelength one would observe if the object is at rest relative to the observer, v_s the velocity of the object, v_r the radial velocity between the source and the observer, and c the velocity of light in the medium. The radiation is red-shifted when its wavelength increases and blue-shifted when its wavelength decreases.

Astronomers use Doppler shifts to derive the movement of stars and other celestial objects toward or away from Earth; for instance, the spectral lines emitted by hydrogen gas in distant galaxies are often observed to be considerably red-shifted. Shifts in frequency are also associated with very strong

© 2015 by Taylor & Francis Group, LLC

gravitational fields, which are referred to as gravitational red-shifts. The cosmological red-shift results from the expansion of space following the Big Bang as well.

The resulting red-shift is commonly expressed in terms of the z-parameter, which is the fractional shift in the spectral wavelength

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_o - \lambda}{\lambda} \tag{1.89}$$

so that

$$\frac{v_r}{c} = \frac{(z+1)^2 - 1}{(z+1)^2 + 1} \approx z \qquad \text{if } z \ll 1 \tag{1.90}$$

In such a situation, one observes the wavelengths stretched out from the receding object and $\Delta\lambda$ becomes a positive number; the spectral lines are shifted to a longer wavelength (a red-shift). On the other hand, wavelengths appear to be compressed from the approaching object and $\Delta\lambda$ turns out to be negative; the spectral lines are shifted to a shorter wavelength (a blue shift).

1.4.4.2 Doppler Broadening

The Doppler broadening, also called the thermal broadening, occurs due to the thermal motion of gas atoms; the distribution of velocities can be found from the Maxwell−Boltzmann distribution. Unlike inhomogeneous broadening where each atom or molecule has a different line-shape, $g(\nu)$, due to its environment, in homogeneous broadening, all atoms behave the same way, i.e., each effectively has the same $g(\nu)$. In the frame of the emitter, a homogeneously broadened line-shape is generated, but in the laboratory frame, seen by an external observer, the homogeneously broadened line-shape is frequency shifted by the Doppler effect, i.e., $\nu_o = \nu_0(1 \pm v/c)$, with ν_o as the observed frequency. The relationship between velocity and frequency is

$$\nu = \nu_0 \left(1 \pm \frac{v}{c}\right) \rightarrow v = c\frac{\nu - \nu_0}{\nu_0} \tag{1.91}$$

and

$$\nu_o = \nu_0 \left(1 \pm \frac{v}{c}\right) \rightarrow d\nu = \frac{\nu_0}{c} dv \rightarrow dv = \frac{c}{\nu_0} d\nu \tag{1.92}$$

therefore, the Doppler broadening is given by

$$g(\nu)d\nu = \sqrt{\frac{mc^2}{2\pi k_B T \nu_0^2}} e^{-\frac{mc^2(\nu-\nu_0)^2}{2k_B T \nu_0^2}} d\nu \tag{1.93}$$

in which $k_B (= 1.38 \times 10^{-23}\ \text{J.K}^{-1})$ represents the Boltzmann constant, m the mass of the atom, and T the absolute temperature of the gas.

To note, the width of a line depends on other factors, for example

- natural broadening that arises from the uncertainty in energy of the states involved in the transition and

© 2015 by Taylor & Francis Group, LLC

- collisional broadening resulting from collisions between atoms, which depends on the frequency of collisions and hence on the density of the gas, and other factors such as Stark broadening and Zeeman broadening.

1.5 Intensity of Waves

The energy density carried by the wave, $V(\mathbf{r}, t)$, is defined as the intensity of the wave and is proportional to $|V(\mathbf{r}, t)|^2$. In other words, the time average of the amount of energy carried by the wave across unit area is perpendicular to the direction of the energy flow in unit time. In many cases, it is proportional to the square of the amplitude of the wave, and for a spherical wave it is inversely proportional to the square of the distance from the source. Electromagnetic theory interprets the light intensity as the energy flux of the field. John Henry Poynting (1852−1914) showed that the intensity is the direct consequence of Maxwell's equations (see Section 2.4). The energy crossing unit area per second is

$$I = \frac{1}{2}\epsilon_0 c\mathbf{E}_0^2 = \frac{1}{2\mu_0}c\mathbf{B}_0^2 \tag{1.94}$$

where $\epsilon_0 (= 8.8541 \times 10^{-12}$ F.m$^{-1})$ and $\mu_0 (= 4\pi k = 4\pi \times 10^{-7}$ H.m$^{-1})$ are the permittivity (or dielectric constant) and the permeability in free space (or vacuum), respectively, c the speed of light in free space, $\mathbf{E}_0$ and $\mathbf{B}_0$, the respective maximum amplitude of the electric and magnetic field vectors; F stands for the farad and H for the henry.

The intensity is defined as the time average of the amount of energy; therefore, the time averaged intensity of the optical field, according to Maxwell's theory is given by

$$I = E_x E_x^* + E_y E_y^* \tag{1.95}$$

The unit of intensity is expressed as the joule per square meter per second (J.m^{-2}s^{-1}) or watt per square meter (W.m^{-2}). The time average is specified as

$$\langle E(t) \rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} E_T(t) dt \tag{1.96}$$

The quantity within the sharp brackets is due to the assumed ergodicity of the field. The word *ergodicity* implies that each ensemble average is equal to the corresponding time average involving a typical member of the ensemble, while the word *stationary* implies that all the ensemble averages are independent of the origin of time. In the monochromatic fields, the electric vector, $\mathbf{E}(\mathbf{r}, t)$, is

© 2015 by Taylor & Francis Group, LLC

expressed in the form

$$\begin{aligned}\mathbf{E}(\mathbf{r},t) &= \Re\left\{\mathbf{A}(\mathbf{r})e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)}\right\} \\ &= \frac{1}{2}\left[\mathbf{A}(\mathbf{r})e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)} + \mathbf{A}^*(\mathbf{r})e^{-i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)}\right] \\ &= \frac{1}{2}\left[\mathbf{A}(\mathbf{r})e^{-i\omega t} + \mathbf{A}^*(\mathbf{r})e^{i\omega t}\right] \end{aligned} \tag{1.97}$$

where the complex vector function $\mathbf{A}(\mathbf{r})$ of position $\mathbf{r} = x, y, z$ is denoted by

$$\mathbf{A}(\mathbf{r}) = a_j(\mathbf{r})e^{i\psi_j(\mathbf{r})} \tag{1.98}$$

in which the amplitudes, $a_j(\mathbf{r})$, and the phases, $\psi_j(\mathbf{r})$, are real functions, $j = 1, 2, 3$, $\boldsymbol{\kappa}$ is the propagation vector, and $\mathbf{A}^*(\mathbf{r})$ represents the complex conjugate of $\mathbf{A}(\mathbf{r})$.

By taking the square of equation (1.97), we obtain

$$\mid \mathbf{E}(\mathbf{r},t) \mid^2 = \frac{1}{4}\left[\mathbf{A}^2(\mathbf{r})e^{-2i\omega t} + \mathbf{A}^{*2}(\mathbf{r})e^{2i\omega t} + 2\mathbf{A}(\mathbf{r})\mathbf{A}^*(\mathbf{r})\right] \tag{1.99}$$

Since the intensities are compared in the same condition, the quantity $\langle|\mathbf{E}(\mathbf{r},t)|^2\rangle$ is used to measure intensity. The average of equation (1.99) can be recast into

$$\begin{aligned}\langle \mid \mathbf{E}(\mathbf{r},t) \mid^2 \rangle &= \frac{1}{4}\left\{\frac{1}{2T}\int_{-T}^{T}\mathbf{A}^2(\mathbf{r})e^{-2i\omega t} + \mathbf{A}^{*2}(\mathbf{r})e^{2i\omega t} + 2\mathbf{A}(\mathbf{r})\mathbf{A}^*(\mathbf{r})dt\right\} \\ &= \frac{1}{4}\left\{\left[\frac{\mathbf{A}^2(\mathbf{r})}{2i\omega T} + \frac{\mathbf{A}^{*2}(\mathbf{r})}{2i\omega T}\right]\left[e^{i\omega T} - e^{-i\omega T}\right] + 2\mathbf{A}(\mathbf{r})\mathbf{A}^*(\mathbf{r})\right\} \end{aligned} \tag{1.100}$$

The response of a detector ([32], [160],[162] and references therein) is governed by the intensity of the optical wave falling on the surface. A detector transforms the light intensity into electrical signal, which provides the number of photo-events collected during the time of measurement and an additive random noise. The detector receives an average of the effects produced by the different values of the amplitude, $a(t)$, which is sensitive to the square of $\mathbf{E}(\mathbf{r},t)$; therefore the time average of the intensity tends to a finite value as the averaging interval is increased indefinitely. The photocurrent at the detector is proportional to the intensity of the light. The average time interval is large compared to the period of the wave in the visible waveband. The intensity of the wave averaged over the time interval $2T$ is needed to make an observation. Therefore, taking the time average of the energy over an interval much greater than the period, $T = 2\pi/\omega$, the intensity I at the same point is derived as

$$\begin{aligned} I &\propto \langle \mid \mathbf{E}(\mathbf{r},t) \mid^2 \rangle = \frac{1}{2}\mathbf{A}(\mathbf{r})\mathbf{A}^*(\mathbf{r}) \\ &= \frac{1}{2}\left(\mid A_x \mid^2 + \mid A_y \mid^2 + \mid A_z \mid^2\right) \\ &= \frac{1}{2}\left(a_1^2 + a_2^2 + a_3^2\right) \end{aligned} \tag{1.101}$$

© 2015 by Taylor & Francis Group, LLC

We note that the first two terms on the RHS of the equation (1.100) are negligible in comparison with the last term, $2\mathbf{A}(\mathbf{r})\mathbf{A}^*(\mathbf{r})$. Assuming a stationary wave field, the intensity of the wave is given by

$$I(\mathbf{r},t) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mid \mathbf{E}(\mathbf{r},t) \mid^2 dt = \langle EE^* \rangle \tag{1.102}$$

The intensity for a stationary wave as defined in equation (1.102) cannot be described adequately for the fluctuations or pulse type phenomena; thus,

$$I(\mathbf{r},t) = \frac{1}{2T} \int_{t-T}^{t+T} \mid \mathbf{E}(\mathbf{r},t) \mid^2 dt \tag{1.103}$$

Equation (1.103) corresponds to the moving average with a window of width T centered at t.

1.6 Interference

The interference effects can be envisaged with devices that combine two or more light waves from the same source at the same time, provided that some conditions are met such as the ability of a wavefront to interfere with a spatially shifted and/or a time-delayed replica of itself. It reveals the correlations (coherence) between light waves; white light interference can also occur since the incoherent light is considered to be the sum of coherent components that interfere. The notion of coherence is defined by the correlation properties between the various quantities of an optical field. The optical coherence is related to the various forms of the correlations of the random processes ([105], [144]). The degree of correlation that exists between the fluctuations in two light waves determines the interference effects arising when the beams are superposed. The correlated fluctuation can be partially or completely coherent. For instance, a polychromatic point source on the sky may produce a fringe packet as a function of an applied path length difference. This fringe packet has an extent referred to as the coherence length, l_c (see Section 8.3). In the case of interferometry, two different polychromatic self-luminous point sources are often considered to be spatially coherent if the fringe packets produced on a detector falling in the same scanning region of the applied path length difference. This type of coherence is known as partial coherence, which applies to extended sources. Partial coherence is a property of two waves whose relative phase undergoes random fluctuations which are not sufficient to make the wave completely incoherent.

An optical interferometer is a device in which beams of light are caused to interfere. The interference manifests itself as alternating regions of constructive (bright) and destructive (dark) interference in the space of relative

© 2015 by Taylor & Francis Group, LLC

propagation delay between the two arms of the interferometer. These alternating bright and dark regions are known as interference fringes. Such an interferometer is designed so that measurements can be made on the interference pattern.

Interferometers based on both two-beam interference and multiple-beam interference of light are powerful tools for metrology and spectroscopy. A wide variety of measurements can be performed, ranging from determining the shape of a surface with an accuracy of a few nanometers to measuring the separation, in millions of kilometers, of binary stars. In spectroscopy, interferometry can be used to determine the hyper-fine structure of spectrum lines. By using lasers in classical interferometers, as well as in holographic interferometers and speckle interferometers, it is possible to perform deformation, vibration, and contour measurements of diffuse objects. There are two basic classes of coherence:

1. Spatial: Spatial coherence describes the correlation between two light waves at different points in space. It determines the surface area that a light source with circular area can illuminate coherently with the light cone of its zeroth diffraction order. This can be improved by increasing the distance to the source; the point sources are more coherent than those with extended source.

2. Temporal: Temporal coherence is determined by the spectral bandwidth and describes the interference of the emitted wave-trains. It compares a light wave with itself at different moments in time, called the coherence time (see Section 8.3), τ_c. A laser beam can have extremely small spectral bandwidth, $\Delta\lambda \sim 10^{-6}\,A$. The narrower the frequency bandwidth, the higher the temporal coherence; the higher the monochromaticity, the longer the coherence time, τ_c.

An astronomical interferometer, for example, Long Baseline Optical Interferometry (LBOI; [33], [137], [161] and references therein) measures spatio-temporal coherence properties of the light emerging from a celestial source. The spatial coherence properties encode the small scale structural content of the intensity distribution in celestial coordinates, while the temporal coherence properties encode the spectral content of the intensity distribution. The measured interferometric signal depends on both structural and spectral content.

1.6.1 Interference of Two Monochromatic Waves

In a homogeneous medium, for regions free of currents and charges, according to either of the equations

$$\begin{aligned} \mathbf{E}(\mathbf{r},t) &= \mathbf{E}_0 e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)} \\ \mathbf{B}(\mathbf{r},t) &= \mathbf{B}_0 e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)} \end{aligned} \tag{1.104}$$

© 2015 by Taylor & Francis Group, LLC

in which $\mathbf{E}_0$ and $\mathbf{B}_0$ are the respective amplitudes of the electric and magnetic field vectors, each rectangular component $V(\mathbf{r},t)$ of the field vectors satisfies the homogeneous wave equation (1.16), at each source-free point.

Let us consider that two monochromatic waves $V(\mathbf{r}_1,t)$ and $V(\mathbf{r}_2,t)$ are superposed at the recombination point, $\mathtt{P}(\mathbf{r})$. The total electric field at this point (output) is

$$V(\mathbf{r},t) = V(\mathbf{r}_1,t) + V(\mathbf{r}_2,t) \tag{1.105}$$

where $V(\mathbf{r},t)$ is given in equation (1.58).

The intensity distribution at the observing screen is given by the sum of the intensities originating from the individual apertures plus the expected value of the cross-product (correlation) of the fields. By squaring equation (1.105), we derive the intensity at point of superposition as

$$\begin{aligned} I(\mathbf{r},t) &= \langle |V(\mathbf{r}_1,t) + V(\mathbf{r}_2,t)|^2 \rangle & (1.106)\\ &= I(\mathbf{r}_1,t) + I(\mathbf{r}_2,t) + 2\Re\left\{\langle V(\mathbf{r}_1,t)V^*(\mathbf{r}_2,t)\rangle\right\} & (1.107)\end{aligned}$$

The third term on the RHS of equation (1.107) is the interference term which depends on the amplitude components, as well as on the phase difference between two waves:

$$\psi_1(\mathbf{r}) - \psi_2(\mathbf{r}) = \delta = \frac{2\pi}{\lambda_0}\Delta\varphi \tag{1.108}$$

Here $\Delta\varphi$ is the optical path difference (OPD) between two waves from the common source to the intersecting point, and λ_0 the wavelength in vacuum. Since the interference term depends on the relative phase, the intensities of superimposed waves cannot be added. The complicated intensity distribution of a diffracted wave is a direct result of the mutual interference of the infinite number of spherical waves emanating from the aperture.

The interference laws of Fresnel−Arago state that (i) two waves linearly polarized in the same plane can interfere, and (ii) two waves linearly polarized with perpendicular polarizations cannot interfere. In the case of the latter, the situation remains the same even if they are derived from perpendicular components of unpolarized light and subsequently brought into the same plane, but they interfere when they are derived from the same linearly polarized wave and subsequently brought into the same plane[115]. Since $\mathbf{A}_1(\mathbf{r})[= a_1(\mathbf{r})e^{i\psi_1(\mathbf{r})}]$ and $\mathbf{A}_2(\mathbf{r})[= a_2(\mathbf{r})e^{i\psi_2(\mathbf{r})}]$ are the complex amplitudes of the two waves, with the corresponding phases ψ_1 and ψ_2, these waves are propagating in the z direction and linearly polarized with the x-component of the electric field vector. Therefore, the total intensity at the same point can be determined as

$$\begin{aligned} I(\mathbf{r},t) &= I(\mathbf{r}_1,t) + I(\mathbf{r}_2,t) + \frac{1}{2}\left[\mathbf{A}(\mathbf{r}_1,t)\cdot\mathbf{A}^*(\mathbf{r}_2,t) + \mathbf{A}^*(\mathbf{r}_1,t)\cdot\mathbf{A}(\mathbf{r}_2,t)\right]\\ &= I(\mathbf{r}_1,t) + I(\mathbf{r}_2,t) + 2\sqrt{I(\mathbf{r}_1,t)I(\mathbf{r}_2,t)}\cos\delta & (1.109)\end{aligned}$$

where $\mathcal{J}_{12} = 2\sqrt{I(\mathbf{r}_1,t)I(\mathbf{r}_2,t)}$ is the interference term.

In general, two light beams are not correlated, but the correlation term,

© 2015 by Taylor & Francis Group, LLC

$\mathbf{A}(\mathbf{r}_1,t)\cdot\mathbf{A}^*(\mathbf{r}_2,t)$, takes on significant values for a short period of time and $\langle\mathbf{A}(\mathbf{r}_1,t)\cdot\mathbf{A}^*(\mathbf{r}_2,t)\rangle = 0$. Time variations of $\mathbf{A}(\mathbf{r})$ are statistical in nature[144]. Hence, one seeks a statistical description of the field (correlations) as the field is due to a partially coherent source.

The interference term enables the positions of fringe intensity maxima and minima to be calculated. The intensity of illumination at $\mathtt{P}(\mathbf{r})$ attains its maximal value when ψ_1 and ψ_2 are in phase. They enhance each other by interfering, and their interference is referred to as constructive interference. The other extreme case occurs when both ψ_1 and ψ_2 are out of phase and tend to cancel each other while interfering. This type of interference is referred to as dark interference. Mathematically we express them as

$$\begin{aligned} I_{max} &= I(\mathbf{r}_1,t) + I(\mathbf{r}_2,t) + 2\sqrt{I(\mathbf{r}_1,t)I(\mathbf{r}_2,t)} \quad \text{if } |\delta| = 0, \pm 2\pi, \cdots \\ I_{min} &= I(\mathbf{r}_1,t) + I(\mathbf{r}_2,t) - 2\sqrt{I(\mathbf{r}_1,t)I(\mathbf{r}_2,t)} \quad \text{if } |\delta| = \pm\pi, \pm 3\pi, \cdots \end{aligned} \tag{1.110}$$

in which I_{max} and I_{min} denote the maximum and minimum intensities in the fringes, and $||$ stands for the absolute.

When $I(\mathbf{r}_1,t) = I(\mathbf{r}_2,t)$, equation (1.109) becomes

$$I = 4I(\mathbf{r}_1,t)\cos^2\frac{\delta}{2} \tag{1.111}$$

The above equation (1.111) reveals that the intensity varies between a maximum value $I_{max} = 4I_1$ and a minimum value $I_{min} = 0$. The intensity at the point of superposition varies between maxima which exceed the sum of the intensities in the beams and minima which may be zero. It provides the information about both the spatial and spectral nature of the source. However, the exact pattern depends on the wavelength of the light, the angular size of the source, and the separation between the apertures or slits. The resolution of an interferometer depends on the separation between the slits and is dictated by the spacing between the maxima, which is known as the fringe angular spacing.

Figure 1.10 depicts the interference of two beams of equal intensity. If one of the apertures is closer to the light source by half of a wavelength, a crest in one light beam corresponds to a trough in the other beam; hence, these two light waves cancel each other, making the source disappear. The light source reappears and disappears every time the delay between the apertures is a multiple of the wave period. If the apertures are kept sufficiently wide, the source is resolved, which means that source can no longer disappear. The actual intensity at any point in the fringe pattern is determined by the amplitude arriving there from each pin hole and the way the two contributions interfere. The contrast or visibility (a dimensionless number between zero and one that indicates the extent to which a source is resolved on the baseline

© 2015 by Taylor & Francis Group, LLC

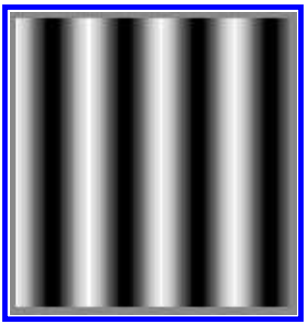


FIGURE 1.10: Interference of the two tilted plane waves gives straight line fringes − the more the tilt the thinner/slimmer the fringes.

being used), $\mathcal{V}$, of the fringe is defined by

$$\begin{aligned} \mathcal{V} &= \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{2\sqrt{I(\mathbf{r}_1, t) I(\mathbf{r}_2, t)}}{I(\mathbf{r}_1, t) + I(\mathbf{r}_2, t)} \\ &= \frac{\text{fringe amplitude}}{\text{average amplitude}} \end{aligned} \tag{1.112}$$

The visibility is the attribute of interference fringes describing the contrast between the bright and dark interference regions. Mathematically, it is a complex (phasor) quantity whose magnitude quantifies the fringe contrast, and phase describes the position of the bright and dark interference relative to some reference. The visibility is related to brightness morphology that an interferometer measures and indicates the extent to which a source is resolved on the baseline used. It contains information about both the spatial and spectral nature of the source. Visibilities can be used to synthesize an image or constrain morphological models of astronomical sources. This is exactly proportional to the amplitude of the image Fourier component corresponding to the fringe spatial frequency, $\mathbf{u}(= u, v)$ vector. It is pertinent to note that astronomical objects are two-dimensional (2D). The 2D space of spatial frequency is called the Fourier plane or the (u, v) coordinates (a plane perpendicular to the source direction) measured in wavelengths. The spatial frequency vector is approximated as

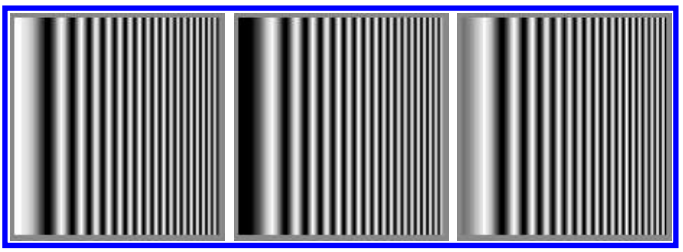
$$\mathbf{u} = \frac{\mathbf{B}}{\lambda} \qquad \text{rad}^{-1}$$

where $\mathbf{B}(= \mathbf{x}_2 - \mathbf{x}_1)$ is the baseline vector with $\mathbf{x}_1(= x_1, y_1)$ and $\mathbf{x}_2(= x_2, y_2)$ the position vectors of the aperture 1 and aperture 2, and λ the wavelength of light. The phase of the fringe pattern is also equal to the Fourier phase of the same spatial frequency component.

When a cylindrical wave interacts with an inclined plane wave, the formula for $\cos(x)^2$ fringes used can be expressed as

$$\mid Ae^{2\pi u x^2} + Be^{2\pi v x} \mid^2 = A^2 + B^2 + 2AB \cos\left[2\pi(ux^2 + vx)\right] \tag{1.113}$$

© 2015 by Taylor & Francis Group, LLC



(a) (b) (c)

FIGURE 1.11: 2D patterns of $\cos^2(2\pi ux^2)$ fringes when a cylindrical wave interacts with a plane wave.

in which the first term of the left hand side (LHS) represents a cylindrical wave, while the second term of the same side represents an inclined plane wave and A and B are the amplitudes of the two waves.

If the plane wave is assumed to be not an inclined one, v turns out to be zero and the nature of the fringe becomes $\cos(2\pi ux^2)$, which can be seen in Figure 1.11. The different light intensities at the side of the three figures are due to the fact that the phase difference between the two interfering waves is different at that point. For example, the first and second patterns (Figures 1.11a and 1.11b) are exactly complementary to each other (implying a phase difference of π at every point), while the third pattern (Figure 1.11c) has a phase shift of approximately $\pi/2$ with respect to the first pattern.

1.6.2 Young's Double-Slit Experiment

The principle of interferometry was first exploited by Thomas Young (1773−1829) at the beginning of the nineteenth century[34]. This experiment, the Young's double-slit experiment, is based on wavefront division which is sensitive to the size and bandwidth of the source. The source consists of a slit illuminated by plane monochromatic light. The light is diffracted by this slit and spreads out on passing through it. The light forms cylindrical wavefronts centered on the slit in accordance with Huygens' principle (see Section 1.7). The diffracted waves fall on an aperture mask in which two slits (see Figure 1.12), $\mathtt{P}_1$ and $\mathtt{P}_2$, are parallel to the source slit and equidistant from it. Thus, the waves at these points have the same phase. They give rise to cylindrical wavefronts similar to that arising at source slit, and the light from each slit spreads over the screen. The light reaching $\mathtt{P}(x, y)$ on a remote opaque screen from $\mathtt{P}_1$ would have a phase dependence on the distance, s_1, depending on the position of $\mathtt{P}(x, y)$ on the screen. Similarly, the light reaching the $\mathtt{P}(x, y)$ from $\mathtt{P}_2$ would have a phase dependence on s_2. The interference pattern is obtained on the remote screen over a plane $x\mathtt{O}y$ normal to a perpendicular bisector $\mathtt{CO}$ of $\mathtt{P}_1\mathtt{P}_2$ and with the x axis parallel to $\mathtt{P}_1\mathtt{P}_2$. Let us assume that a is the distance between the aperture mask and interference point at $\mathtt{P}$, where

© 2015 by Taylor & Francis Group, LLC

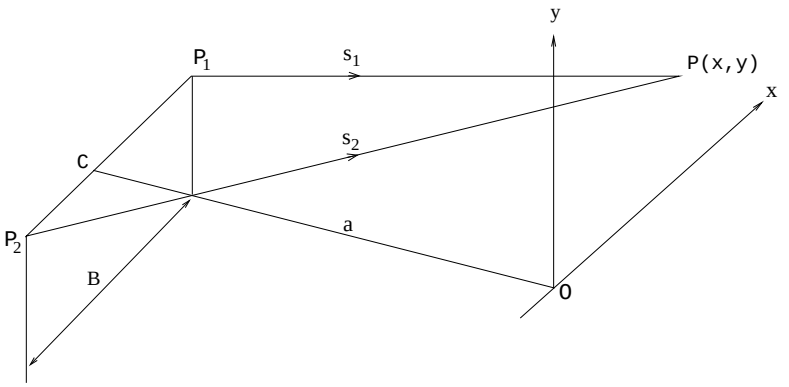


FIGURE 1.12: Illustration of interference with two point sources.

$a \gg B$,

$$\begin{aligned} s_1 = \mathtt{P_1P} &= \left[a^2 + y^2 + \left(x - \frac{B}{2}\right)^2\right]^{1/2} \\ s_2 = \mathtt{P_2P} &= \left[a^2 + y^2 + \left(x + \frac{B}{2}\right)^2\right]^{1/2} \end{aligned} \tag{1.114}$$

where B is the separation (baseline) of the pinholes and is assumed to be an order of magnitude λ.

The net effect at $\mathtt{P}(x, y)$ depends on the difference of s_1 and s_2. If this difference is an integral number of wavelengths, the waves reinforce and the screen is bright. If the difference is half a wavelength more than an integral number, the waves interfere destructively and the screen is dark. By squaring these sub equations (1.114) followed by subtracting, we obtain

$$s_2^2 - s_1^2 = 2xB \tag{1.115}$$

The geometrical path difference between the spherical waves reaching the observation point, $\mathtt{P}$, is caused by the difference of propagation distances of the waves from the pinholes, $\mathtt{P}_2$ and $\mathtt{P}_1$, to $\mathtt{P}$ and is expressed as

$$\Delta s = \frac{xB}{a} = B \sin\theta \tag{1.116}$$

in which θ is the angle OCP. The observed intensity along the observation screen is given by

$$I = I_{max} \cos^2\left(\frac{\kappa B \sin\theta}{2}\right) \tag{1.117}$$

The phase difference, δ, resulting from the difference in propagation distance is of the form

$$\delta = \kappa \Delta\varphi = \kappa B \sin\theta = 2\pi\left(\frac{B \sin\theta}{\lambda}\right) \tag{1.118}$$

© 2015 by Taylor & Francis Group, LLC

If n is the refractive index of the homogeneous medium, the different optical path from $\mathtt{P}_1$ and $\mathtt{P}_2$ to the point $\mathtt{P}$, the optical path difference (OPD), $\Delta\varphi$, is given by

$$\Delta\varphi = n\Delta s = \frac{nxB}{a} \tag{1.119}$$

and the corresponding phase difference is

$$\delta = \frac{2\pi}{\lambda_0}\frac{nxB}{a} \tag{1.120}$$

Other phase differences arising as a result of propagation through different media or initial phase differences are required to be summed up into a total phase difference $\delta(\theta)$. Thus, the intensity observed at $\mathtt{P}$ is derived as

$$I = I_{max}\cos^2\left(\frac{\delta}{2}\right) \tag{1.121}$$

The angle $\mathtt{P}_1\mathtt{P}\mathtt{P}_2$ is considered to be very small, so that the intensity can be derived from equation (1.109). Waves are in phase interfere constructively at a certain point if the distance traveled by one wave is the same or if it differs by an integral number of wavelengths from the path length traveled by the second wave. For the waves to interfere destructively, the path lengths must differ by an integral number of wavelengths plus half a wavelength. According to equations (1.110) and (1.120), there are maxima and minima of intensities, respectively, when

$$x = \begin{cases} ma\lambda_0/nB & \mid m \mid = 0, 1, 2, \cdots \\ ma\lambda_0/nB & \mid m \mid = 1/2, 3/2, 5/2, \cdots \end{cases} \tag{1.122}$$

The interference patterns in the immediate vicinity of $\mathtt{O}$ are equidistant and are at right angles to the line $\mathtt{P}_1\mathtt{P}_2$ joining the two sources. The separation of adjacent bright fringes, Δ, is proportional to the wavelength, λ_0 and inversely proportional to the baseline between the apertures B, that is,

$$\Delta = \frac{a\lambda_0}{nB} \tag{1.123}$$

The order of the interference at any point is given by

$$m = \frac{\delta}{2\pi} = \frac{\Delta\varphi}{\lambda_0} \tag{1.124}$$

1.6.3 Michelson Interferometer

The Michelson classical interferometer is directly related to the monochromaticity of radiation. It is based on amplitude division and generally used to measure the temporal coherence of a source. Light from a source, $\mathtt{S}$, is

© 2015 by Taylor & Francis Group, LLC

collimated by a lens, $\mathtt{L}_1$, and is partly reflected, partly transmitted, by the half-silvered mirror called a beam-splitter, BS, (see Figure 1.13). The two beams leaving BS are reflected by the mirrors, $\mathtt{M}_1$ (kept at a movable arm) and $\mathtt{M}_2$ (kept at a fixed arm). The beam from $\mathtt{M}_2$ is partly transmitted and partly reflected at BS and the transmitted part propagates to a screen, O, through a lens, $\mathtt{L}_2$. Similarly, the beam from $\mathtt{M}_1$ is partly reflected to the screen. The extra path traversed by one of the beams is compensated by translating $\mathtt{M}_1$. Successive maxima and minima of the fringes are observed at the output with a periodicity governed by the ratio of the optical path difference (OPD) to the wavelength. All the wavelengths add in phase at zero OPD. The loss of visibility away from zero OPD refers to the blurring due to stretched interference fringes. This stresses the importance of the spectrum of the source.

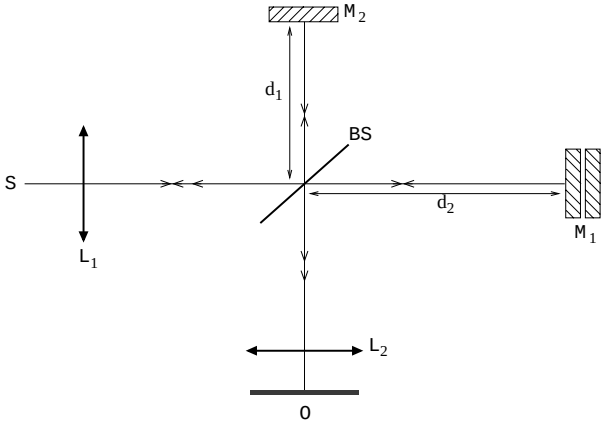


FIGURE 1.13: Schematic diagram of a classical Michelson interferometer.

With the interferometer each monochromatic component produces an interference pattern as the path difference is increased from zero; two component patterns show increasing mutual displacement, because of the difference of wavelength. The visibility of the fringes therefore decreases and they disappear when the OPD is sufficiently large. The angular dimension of the source producing fringes can be determined simply by observing the smallest value of d for which the visibility of the fringes is minimum. This condition occurs when

$$d = \frac{A\lambda_0}{\theta} \tag{1.125}$$

where $A = 0.5$ for two point sources of angular separation θ, and $A = 1.22$ for a uniform circular disc source of angular diameter θ.

The Michelson classical interferometer, which revealed that the speed of light is independent of the observer's velocity, was used for the Michelson−Morley[35] experiment to detect the relative motion of matter through the stationary luminiferous aether, for which, A. Michelson was

© 2015 by Taylor & Francis Group, LLC

awarded the Nobel prize in 1907. Such an interferometer can also be used to measure refractive indices very accurately.

1.7 Diffraction

Due to the scattering of the light by inhomogeneities in transparent media, aberrations of the optical system, diffraction and nonideal images are formed. Loss of rays by such means diminishes the brightness of the image. Imaging, in general terms, means a reproduction/production of desired characteristic features of a 3D object (real or simulated). Every optical system intercepts a portion of the wavefront emerging from the object; as a result the image cannot be sharp. Also the amplitudes along the wavefront are nonuniform and the propagation is not linear. The effect of using a limited portion of the wavefront leads to diffraction that has played a major role in elaborating the theory of light since the description of its effects in the 17th century. The operative theory of diffraction by Auguste Jean Fresnel (1788−1827) based on the principle developed by Huygens and the theory of interference developed by Young (see Section 1.6.2), followed by the experimental confirmation of some unexpected predictions, marks the importance of this new theory. A firm mathematical foundation of the theory was obtained later by Kirchhoff[105].

1.7.1 Huygens−Fresnel Principle

As stated earlier in Section 1.6, the interference phenomena are produced by the superposition of beams of light which come from a single source by different paths. The diffraction occurs when some obstacle blots out part of a wavefront. Fresnel added the principle of interference into the Huygens principle, which states that

- each element of the wavefront at a point may be considered as the origin of a secondary disturbance, which gives rise to spherical wavelets, a series of tiny spherical waves which repeatedly generate themselves at all points across the wavefront and propagate outwards,
- the position of wavefront at any later time is an envelope of such wavelets,
- frequency and speed of wavelets are the same as those of primary, and
- the secondary wavelets radiating from all points in that position of the wavefront which is not blocked by the obstacle, mutually interfere; the amplitude at any later point is the superposition of wavelets.

© 2015 by Taylor & Francis Group, LLC

The diffraction effect disappears when the wavelength approaches zero. In general, diffraction pattern in the plane of observation can be derived in terms of the Huygens−Fresnel principle, which is an integral of the contribution from wavelets coming from an aperture, $\mathtt{W}$, which is regarded as having been formed by taking all the Fresnel zones, extending to infinity, and blotting out all but a limited number at the center. Conversely, a circular obstacle is considered as having been formed by blotting out a limited number of zones at the center, leaving all the rest, out to infinity, to make their contributions to the intensity at $\mathtt{P}(\mathbf{r})$, where $\mathbf{r}$ is the position vector of a point (x, y, z).

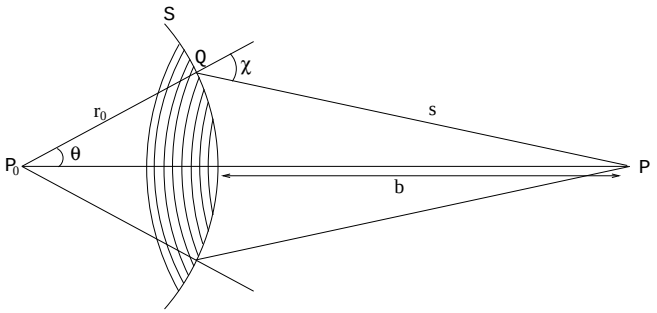


FIGURE 1.14: Construction of the Fresnel zone.

Let us consider that a monochromatic wave emitted from a point source $\mathtt{P}_0(\mathbf{r}_0)$ falls on an aperture $\mathtt{W}$, and $\mathtt{S}$ is the instantaneous position of a spherical monochromatic wavefront of radius r_0 (see Figure 1.14). At $\mathtt{Q}(\mathbf{r}')$, the disturbance is represented by $Ae^{-i\kappa r_0}/r_0$, in which A is the amplitude at unit distance from the source and κ the wavenumber; here the time periodic factor, $e^{-i\omega t}$ is omitted. The expression for the elementary contribution, $dU(\mathbf{r})$, of an area, $d\mathtt{S}$, at the point, $\mathtt{Q}(\mathbf{r}')$, is

$$dU(\mathbf{r}) = K(\chi)\frac{Ae^{i\kappa r_0}}{r_0}\frac{e^{i\kappa s}}{s}d\mathtt{S}$$

where s is the distance between the points $\mathtt{Q}(\mathbf{r}')$ and $\mathtt{P}(\mathbf{r})$, $K(\chi)$ the scaling factor which accounts for the properties of the secondary wavelet, χ the angle of diffraction between the normal at $\mathtt{Q}(\mathbf{r}')$ and the direction $\mathtt{Q}(\mathbf{r})\ \mathtt{P}(\mathbf{r})$.

The scaling factor depends on the angle, χ, being maximum for $\chi = 0$ and zero for $\chi = \pi/2$. The total disturbance at the point, $\mathtt{P}(\mathbf{r})$, is deduced as[105].

$$U(\mathbf{r}) = \frac{Ae^{i\kappa r_0}}{r_0}\int_{\mathtt{W}}\frac{e^{i\kappa s}}{s}K(\chi)d\mathtt{S} \tag{1.126}$$

Fresnel evaluated the integral by considering in the diffraction aperture successive zones of constant phase, i.e., for which the distance s is constant within $\lambda/2$. The field at $\mathtt{P}(\mathbf{r})$ yields from the interference of the contributions of these zones. According to Fresnel, the scaling factor is $K(\chi) = -i(1 + \cos\chi)/(2\lambda)$.

© 2015 by Taylor & Francis Group, LLC

1.7.2 Kirchhoff's Scalar Diffraction Theory

A rigorous theory on the diffraction of light can be obtained from a wave equation derived from Maxwell's equations (see Section 2.4) with suitable conditions. If one neglects the coupling between electric and magnetic fields and treats light as a scalar field, decisive simplifications arise. The scalar theory yields correct values under two conditions: (i) the diffracting aperture must be large compared with a wavelength and (ii) the diffracting fields must not be observed too close to the aperture.

Kirchhoff[36] showed that the Huygens−Fresnel principle may be regarded as an approximate form of a certain integral theorem, which expresses the solution of the homogeneous wave equation, at a point in the field, in terms of the values of the solution and its first derivatives at all points on a surface surrounding $\mathtt{P}(\mathbf{r})$. However, his theory applies to the diffraction of scalar waves. Let us consider a monochromatic scalar wave equation[105]:

$$V(\mathbf{r},t) = U(\mathbf{r})e^{-i\omega t} \tag{1.127}$$

where the complex function $U(\mathbf{r})$ must satisfy the time-independent wave equation

$$\left(\boldsymbol{\nabla}^2 + \kappa^2\right) U = 0 \tag{1.128}$$

with $\boldsymbol{\nabla}^2$ as Laplacian operator (see equation B.14; Appendix B.1.4).

Named after Hermann von Helmholtz (1821−1894), equation (1.128) is known as the Helmholtz equation. In order to obtain a solution of V in all the space within a closed volume $\mathtt{V}$ with a boundary surface $\mathtt{S}$ in terms of source function and the given value of V, let us consider that U' is another complex valued function which satisfies the same continuity requirements as U. If these functions, U and U' and their first and second partial derivatives are single valued and continuous within and on $\mathtt{S}$, by applying the Green's theorem (see Appendix B.4), we get

$$\int_{\mathtt{V}} \left(U\boldsymbol{\nabla}^2 U' - U'\boldsymbol{\nabla}^2 U\right) d\mathtt{V} = \int_{\mathtt{S}} \left(U\frac{\partial U'}{\partial n} - U'\frac{\partial U}{\partial n}\right) d\mathtt{S} \tag{1.129}$$

where $\partial/\partial n$ denotes a partial derivative along the inward normal to $\mathtt{S}$.

1.7.2.1 Kirchhoff's Diffraction Integral

Within the volume, $\mathtt{V}$, the disturbance U' is an expanding spherical wave; hence, it satisfies the Helmholtz equation,

$$\left(\boldsymbol{\nabla}^2 + \kappa^2\right) U' = 0 \tag{1.130}$$

The value of U' at any arbitrary point $\mathtt{P}(\mathbf{r})$ is of the form

$$U'(\mathbf{r}_1) = \frac{e^{i\kappa s}}{s} \tag{1.131}$$

© 2015 by Taylor & Francis Group, LLC

Here we adopt the notation that s is the length of the vector $\mathbf{s}$ pointing from $\mathtt{P}(\mathbf{r})$ to $\mathtt{P}_1(\mathbf{r}_1)$. Green's function U' has a singularity at $s = 0$. The singularity at $\mathtt{P}(\mathbf{r})$ is avoided by surrounding $\mathtt{P}(\mathbf{r})$ with a spherical surface $\mathtt{S}'$ of a small radius, ε (see Figure 1.15). The enclosed region $\mathtt{V}(= 4\pi\varepsilon^3/3)$ is lying between $\mathtt{S}$ and $\mathtt{S}'$. By substituting the two Helmholtz equations (1.128) and (1.130) in to equation (1.129), we find

$$\begin{aligned}\int_{\mathtt{V}} \left(U\boldsymbol{\nabla}^2 U' - U'\boldsymbol{\nabla}^2 U\right) d\mathtt{V} &= -\int_{\mathtt{V}} \left(UU'\kappa^2 - U'U\kappa^2\right) d\mathtt{V} \\ &= \int_{\mathtt{S}} \left(U\frac{\partial U'}{\partial n} - U'\frac{\partial U}{\partial n}\right) d\mathtt{S} = 0 \quad (1.132)\end{aligned}$$

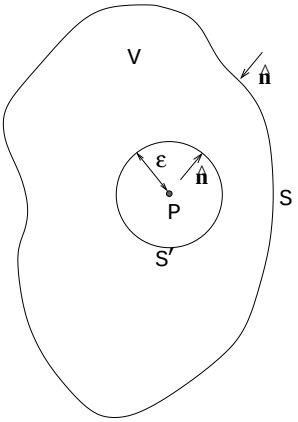


FIGURE 1.15: Two closed surfaces around a point, P, where a spherical wavelet has a singularity.

Let us extend the integration throughout the volume between $\mathtt{S}$ and the surface $\mathtt{S}'$. By invoking equation (1.131) and replacing the value of U' in equation (1.132), we obtain

$$\int_{\mathtt{S}} \left[U\frac{\partial}{\partial n}\left(\frac{e^{i\kappa s}}{s}\right) - \frac{e^{i\kappa s}}{s}\frac{\partial U}{\partial n}\right] d\mathtt{S} + \int_{\mathtt{S}'} \left[U\frac{\partial}{\partial n}\left(\frac{e^{i\kappa s}}{s}\right) - \frac{e^{i\kappa s}}{s}\frac{\partial U}{\partial n}\right] d\mathtt{S} = 0 \quad (1.133)$$

It is reasonable to take the values of $U = Ae^{i\kappa s}/s$, in which A is the amplitude. The partial derivative of U in the outward normal direction at the surface $\mathtt{S}$ is given by

$$\begin{aligned}\frac{\partial U}{\partial n} &= \frac{Ae^{i\kappa s}}{s}\left[i\kappa - \frac{1}{s}\right]\cos(\hat{\mathbf{n}}, \mathbf{s}) \\ &\approx \frac{Ae^{i\kappa s}}{s} i\kappa \cos(\hat{\mathbf{n}}, \mathbf{s}) \quad (1.134)\end{aligned}$$

© 2015 by Taylor & Francis Group, LLC

where the approximate relation holds if $s \gg \lambda$ and $(\hat{\mathbf{n}}, \mathbf{s})$ represents the cosine of the angle between the normal $\hat{\mathbf{n}}$ and the vector $\mathbf{s}$, joining $\mathtt{P}(\mathbf{r})$ to $\mathtt{P}_1(\mathbf{r}_1)$.

When the surface $\mathtt{S}'$ is considered, the normal n is directed towards point $\mathtt{P}(\mathbf{r})$; therefore, $\cos(\hat{\mathbf{n}}, \boldsymbol{\varepsilon}) = -1$. Since ε is very small, by invoking equation (1.134), the partial derivative $\partial U'/\partial n$ on the surface is simplified as

$$\frac{\partial U'}{\partial n} = \frac{e^{i\kappa\varepsilon}}{\varepsilon}\left(\frac{1}{\varepsilon} - i\kappa\right) \tag{1.135}$$

By substituting equations (1.134 and 1.135) in to equation (1.133), we get

$$\int_{\mathtt{S}} \frac{e^{i\kappa s}}{s}\left[\frac{\partial U}{\partial n} - Ui\kappa\cos(\hat{\mathbf{n}}, \mathbf{s})\right] d\mathtt{S} + \int_{\mathtt{S}'} \frac{e^{i\kappa\varepsilon}}{\varepsilon}\left[\frac{\partial U}{\partial n} - U\left(\frac{1}{\varepsilon} - i\kappa\right)\right] d\mathtt{S}' = 0 \tag{1.136}$$

Since the integral over $\mathtt{S}$ is independent of ε, as $\varepsilon \to 0$, the second integral is replaced by its limiting value. Thus, for $\varepsilon \to 0$, and assuming continuity of U and its derivative, the second integral in equation (1.136) takes the form

$$\begin{aligned}\int_{\mathtt{S}'} \frac{e^{i\kappa\varepsilon}}{\varepsilon}\left[\frac{\partial U}{\partial n} - U\left(\frac{1}{\varepsilon} - i\kappa\right)\right] d\mathtt{S}' &= 4\pi\varepsilon^2 \frac{e^{i\kappa\varepsilon}}{\varepsilon}\left[\frac{\partial U}{\partial n} - U\left(\frac{1}{\varepsilon} - i\kappa\right)\right] \\ &= -4\pi U(\mathbf{r}) \end{aligned} \tag{1.137}$$

where $U_{\varepsilon\to 0} = U(\mathbf{r})$.

Hence, from equations (1.136) and (1.137)

$$U(\mathbf{r}) = \frac{1}{4\pi}\int_{\mathtt{S}}\left\{\frac{e^{i\kappa s}}{s}\frac{\partial U}{\partial n} - U\frac{\partial}{\partial n}\left(\frac{e^{i\kappa s}}{s}\right)\right\} d\mathtt{S} \tag{1.138}$$

This result (equation 1.138) is known as the Helmholtz−Kirchhoff integral theorem. It is important to note that as $\kappa \to 0$, the time-independent wave equation (1.128) reduces to the Laplace equation, $\boldsymbol{\nabla}^2 U = 0$, and equation (1.138) goes over into the formula of potential theory

$$U(\mathbf{r}) = \frac{1}{4\pi}\int_{\mathtt{S}}\left\{\frac{1}{s}\frac{\partial U}{\partial n} - U\frac{\partial}{\partial n}\left(\frac{1}{s}\right)\right\} d\mathtt{S} \tag{1.139}$$

The potential theory states that if a 2D potential function and its normal derivative vanish together along any finite curve segment, the potential function must vanish over the entire plane. This is true for solution of a 3D wave equation (see equation 1.16) as well. In order to derive the general form of Kirchhoff's theorem which applies to nonmonochromatic waves, let us consider that $V(\mathbf{r}, t)$ is a solution of the said wave equation, and V is represented in the form of a Fourier integral:

$$V(\mathbf{r}, t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \widehat{U}(\mathbf{r}) e^{-i\omega t} d\omega \tag{1.140}$$

Since $V(\mathbf{r}, t)$ is assumed to satisfy the wave equation (1.16), $\widehat{U}(\mathbf{r})$ satisfies

© 2015 by Taylor & Francis Group, LLC

the time-independent equation (1.128). If V obeys the regularity conditions within and on a closed surface $\mathtt{S}$, then using Kirchhoff's formula separately to each Fourier component

$$U(\mathbf{r}) = -\frac{1}{4\pi}\int_{\mathtt{S}}\left\{\widehat{U}\frac{\partial}{\partial n}\left(\frac{e^{i\kappa s}}{s}\right) - \frac{e^{i\kappa s}}{s}\frac{\partial\widehat{U}}{\partial n}\right\}d\mathtt{S} \tag{1.141}$$

On using equation (1.140), the general form of Kirchhoff's theorem can be written as[105]

$$V(\mathbf{r},t) = -\frac{1}{4\pi}\int_{\mathtt{S}}\left\{[V]\frac{\partial}{\partial n}\left(\frac{1}{s}\right) - \frac{1}{cs}\frac{\partial s}{\partial n}\left[\frac{\partial V}{\partial t}\right] - \frac{1}{s}\left[\frac{\partial V}{\partial n}\right]\right\}d\mathtt{S} \tag{1.142}$$

The square brackets denote values of the functions taken at the time $t - s/c$, known as retarded values. The first two terms on the RHS of equation (1.142) represent a contribution of doublets of strength $V/4\pi$ per unit area. while the last term represents the contribution of a distribution of sources of strength $-\frac{1}{4\pi}\frac{\partial V}{\partial n}$ per unit area.

1.7.2.2 Kirchhoff's Boundary Conditions

As stated earlier in Section 1.7.1, the diffraction effects result from the obstruction of a wavefront of light. If the medium in which the waves propagate is of finite extension and is bounded, the boundaries of that medium affect the propagating waves. The waves bounce against these boundaries and partly reflect back into the medium and partly transmit outside the medium. The reflected waves superimpose on incident waves, which may, in turn, cancel each other. If the volume of the medium is finite, the propagation vector, $\boldsymbol{\kappa}$, becomes one of a discrete set of values.

Another boundary effect is the diffraction of waves from apertures and opaque screens. Let a plane wave of frequency ω and wave vector $\boldsymbol{\kappa}$ be incident on an infinite opaque screen containing a finite aperture; the wave decays at infinity on the other side of the opaque plane. These data form a well-defined boundary value problem for the wave equation (1.16). The boundary values define the wave at any point and at any time on the rear side of the opaque screen. The diffracted wave on the other side of the screen is constructed by considering each point of the aperture of the opaque screen as a point source of waves, as well as of same ω, and then, superposing these waves on the other side of the screen. Point sources for waves obeying the said wave equation generate spherical wavefronts; the form of these spherical waves in three-space dimensions is represented by equation (1.25). The total sum over all the emerging spherical waves at a point on the observation plane provides the diffracted wave at that point.

Let us consider a monochromatic wave, from a point source, is propagating through an opening in a plane opaque screen with aperture, $\mathtt{W}$ (see Figure 1.16). In order to apply equation (1.138), a closed surface, $\mathtt{S}$, around a point, $\mathtt{P}$, can

© 2015 by Taylor & Francis Group, LLC

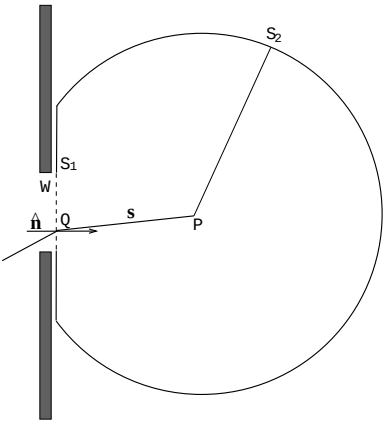


FIGURE 1.16: Diffraction by an arbitrary aperture, $\mathtt{W}$.

be chosen in a manner that $\mathtt{S} = \mathtt{W} + \mathtt{S}_1 + \mathtt{S}_2$, in which $\mathtt{S}_1$ is the area behind the opaque screen and $\mathtt{S}_2$ the large surface. To evaluate the solution for $U(\mathbf{r})$ in this case, some assumptions on the value of U and its first derivative on the surface $\mathtt{S}$ are needed. Kirchhoff's boundary conditions are

- the field distribution U and its derivative $\partial U/\partial n$ within the aperture are the same as they would be in the absence of the screen and
- in an other area of the screen, i.e., on $\mathtt{S}_1$, that lies in the geometrical shadow of the screen, the field distribution U and its derivative $\partial U/\partial n$ are

$$U = 0 \qquad \frac{\partial U}{\partial n} = 0 \tag{1.143}$$

1.7.3 Fresnel–Kirchhoff Diffraction Formula

In order to study diffraction at a plane opaque screen with a opening $\mathtt{W}$ (see Figure 1.17), let us consider a monochromatic wave, from a point source $\mathtt{P}_0(\mathbf{r}_0)$ to propagate through the said opening.

The linear dimensions of the opening are assumed to be small compared to the distances of the screen from the point $\mathtt{P}_0(\mathbf{r}_0)$ and the point, $\mathtt{P}(\mathbf{r})$, at which light disturbance is to be determined. The dimension of the opening is large compared to the wavelength of light. The disturbance at $\mathtt{P}(\mathbf{r})$ is obtained by using Kirchhoff's integral and boundary conditions, as

$$U(\mathbf{r}) = -\frac{1}{4\pi}\int_{\mathtt{W}}\left[U\frac{\partial}{\partial n}\left(\frac{e^{i\kappa s}}{s}\right) - \frac{e^{i\kappa s}}{s}\frac{\partial U}{\partial n}\right]d\mathtt{S} \tag{1.144}$$

where s is the distance of the element $d\mathtt{S}$ from $\mathtt{P}(\mathbf{r})$, $\mathbf{r}$ is the position vector

© 2015 by Taylor & Francis Group, LLC

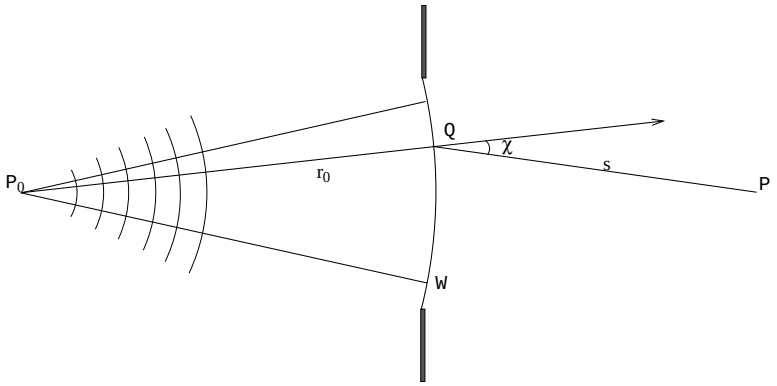


FIGURE 1.17: Contribution from spherical wavelets from an aperture in terms of Huygens−Fresnel principle.

and

$$U = \frac{Ae^{i\kappa r}}{r} \tag{1.145}$$

is the field at any point in the opening $\mathtt{W}$ in which A is the amplitude of the incident wave at unit distance from the source.

The term $e^{i\kappa r}/r$ is the field due to a point source radiating isotropically. Neglecting the normal derivatives, the terms $1/r$ and $1/s$ in comparison to κ, we obtain from equation (1.144)

$$U(\mathbf{r}) = \frac{A}{i\lambda}\int_{\mathtt{W}} \frac{e^{i\kappa(r+s)}}{rs}\left[\frac{\cos(\hat{\mathbf{n}},\mathbf{s}) - \cos(\hat{\mathbf{n}},\mathbf{r})}{2}\right] d\mathtt{S} \tag{1.146}$$

where one of Kirchhoff's boundary conditions has been used.

Equation (1.146) is known as the Fresnel−Kirchhoff diffraction formula implying the reciprocity theorem, which states that a point source at point $\mathtt{P}_0$ would produce a point $\mathtt{P}$, the same effect that a point source of equal intensity placed at $\mathtt{P}$ would produce at the point $\mathtt{P}_0$. Also, this equation includes obliquity factor, $K(\chi)$, as well as a phase-shift, $e^{-i\pi/2}$, that is accounted by

- the secondary wave oscillating a quarter of a wave out of phase with the primary and
- the amplitudes of the secondary wave and that of the primary wave being in the ratio $1 : \lambda$.

If the radius of curvature of the wave is sufficiently large, $\cos(\hat{\mathbf{n}}, \mathbf{r}_0) = 0$ on $\mathtt{W}$. Let us set $\chi = \pi - (\mathbf{r}_0, \mathbf{s})$, so that

$$U(\mathbf{r}) = \frac{Ae^{i\kappa r_0}}{2i\lambda r_0}\int_{\mathtt{W}} \frac{e^{i\kappa s}}{s}(1+\cos\chi)d\mathtt{S} \tag{1.147}$$

© 2015 by Taylor & Francis Group, LLC

where r_0 is the radius of the wavefront $\mathtt{W}$ and χ is the angle of diffraction between the normal at Q and the direction $\mathtt{QP}$.

When the points $\mathtt{P}_0(\mathbf{r}_0)$ and $\mathtt{P}(\mathbf{r})$ are very near to the axis, $\cos(\hat{\mathbf{n}}, \mathbf{s}) = 1$ and $\cos(\hat{\mathbf{n}}, \mathbf{r}) = -1$; thus, the equation (1.146) becomes

$$U(\mathbf{r}) = \frac{A}{i\lambda} \int_{\mathtt{W}} \frac{e^{i\kappa(r+s)}}{rs} d\mathtt{S} \tag{1.148}$$

If the Babinet principle holds[37], then for two complimentary screens (the openings in one correspond exactly to the opaque portions of the other and vice versa, i.e., the openings in the two screens add up to fill the whole plane), we have

$$U_1 + U_2 = U \tag{1.149}$$

in which $U_1(\mathbf{r})$ and $U_2(\mathbf{r})$, respectively, are the values of the complex displacement at the point of observation when the first or the second screen is placed between the source and the point of observation, and U is the value of the distance when no screen is present.

If $U = 0$, then $U_1 = -U_2$, which implies that at points where U is zero, the phases of U_1 and U_2 differ by π, and the intensities $I_1 = |U_1|^2$ and $I_2 = |U_2|^2$ are equal. If $U_1 = 0$, then $U = U_2$, i.e., at points where the intensity is null in the presence of one of the screens, the intensity in the presence of the other is the same as if the screen were absent.

1.7.4 Rayleigh−Sommerfeld Integral

Kirchhoff's boundary conditions (see Section 1.7.2.2) imposed on the diffraction problem in order to obtain the Fresnel−Kirchhoff solution are not always consistent[127]. For instance, if $U = \partial U/\partial n \equiv 0$ on $\mathtt{S}_1$, as has been assumed in the said boundary condition, the solution to the wave equation becomes zero everywhere, i.e., $U \equiv 0$ over the entire space. The diffraction formula given by equation (1.146) fails to reproduce the assumed boundary condition as well. In 1896, Sommerfeld modified Kirchhoff's law using the theory of Green's function, the result of which is the Rayleigh−Sommerfeld diffraction theory. He eliminated the need of imposing boundary values on the disturbance and its derivative simultaneously. In what follows, we describe the Rayleigh−Sommerfeld diffraction integral for U at distance z behind an aperture in a planar mask, which is given by

$$U(\mathbf{r}) = \frac{Ae^{i\kappa r_0}}{i\lambda r_0} \int_{\mathtt{W}} \frac{e^{i\kappa s}}{s} \cos\chi d\mathtt{S} \tag{1.150}$$

with $\cos\chi = \frac{z}{s}$, s as the distance from a point in the aperture to a point on the screen, and χ the angle the diffracted rays make with the z axis.

Equation (1.150) can be recast as

$$U(\mathbf{r}) = \frac{1}{i\lambda} \int_{-\infty}^{\infty} U(\xi, \eta). \frac{e^{i\kappa s}}{s} \cos\chi d\mathtt{S} \tag{1.151}$$

© 2015 by Taylor & Francis Group, LLC

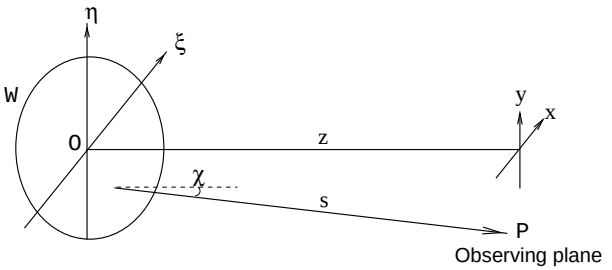


FIGURE 1.18: Plane geometry to describe diffraction problem.

in which (ξ, η) are the coordinates of a point on the input plane, $\mathbf{r}(= x, y, z)$ the coordinates of a point on the observing plane, and

$$U(\xi,\eta) = \begin{cases} Ae^{i\kappa r_0}/r_0 & \text{if } (\xi,\eta) \in \mathtt{W} \\ 0 & \text{otherwise} \end{cases} \tag{1.152}$$

In Figure 1.18, we have considered only small diffraction angles, so that $K(\chi) \approx 1$. The size of the diffracting aperture and the region of observation of the diffracted rays must be small compared to their distance, z, so that $1/s \approx 1/z$. However, λ is very small compared to s and z, and it is not probable to approximate $e^{i\kappa s}$ by $e^{i\kappa z}$. Thus,

$$U(x,y) = \frac{1}{i\lambda z}\int_{-\infty}^{\infty} U(\xi,\eta)e^{i\kappa s}d\xi d\eta \tag{1.153}$$

with

$$\begin{aligned} s &= \sqrt{z^2 + (x-\xi)^2 + (y-\eta)^2} \\ &= z\sqrt{1 + \left(\frac{x-\xi}{z}\right)^2 + \left(\frac{y-\eta}{z}\right)^2} \end{aligned} \tag{1.154}$$

We have noticed thus far that a correct result for diffraction problem depends on the specification of the field on the boundary, i.e., on the diffraction aperture. In what follows, we discuss briefly two more approximations:

1. Debye approximation: In the case of focusing with high numerical aperture, one uses the Debye theory. The field in the focal region is a superposition of plane waves whose propagation vector falls inside the geometrical cone formed by drawing straight lines from the focal point through the edge of the aperture.

2. Paraxial approximation: A wave is said to be paraxial if it gets displaced differentially from the optical axis. One may use the first order approximation to the trigonometric functions in the equations for a differential

© 2015 by Taylor & Francis Group, LLC

traced ray. Hence, the resulting paraxial ray equations are identical to the first order equations.

Most diffraction problems occur when a light wave propagates along a direction which is closed to the axis of optical components, such as lenses or apertures, where the transformation is imperfect. A homocentric bundle of rays originating from an object point is, in general, transformed into a heterocentric bundle of rays in the image space. In terms of wavefronts, the spherical wavefronts emerging from the object point are transformed into aspherical wavefronts. In some other situations, even though there is a homocentric bundle or spherical wavefronts in the image space, neither the center of the bundle nor the center of the spherical wavefronts is located at the desired image point corresponding to the object point. This departure from ideal image formation gives rise to a loss of sharpness and resolution in the images formed by real optical imaging systems. This deviation in functiotning of real optical imaging systems from ideal systems is termed aberration, knowledge of which is of paramount importance in analysis and synthesis of optical imaging systems. The aberrations in geometrical optics cause different rays to converge at different points. These aberrations generate from subsets of a larger centered bundle of spherically aberrated rays. Typically, an optical system suffers from spherical aberration, coma, astigmatism, and chromatic aberration.

In such a case, paraxial approximation may be assumed, under which a homocentric bundle of rays emerging from an object is transformed into another homocentric bundle of rays converging to the image point corresponding to the object; equivalently, spherical wavefronts emerging from the point object are transformed into perfectly spherical wavefronts with their common center at the corresponding image point in the image space[38]. Of course, on account of the finite size of the pupils, a part of the sphere takes part in actual image formation.

1.7.5 Fresnel Approximation

The Fresnel approximation yields good results for the near field diffraction region which begins at some distance from the aperture; hence, the curvature of the wavefront must be taken into account. Both the shape and the size of the diffraction pattern depend on the distance between aperture and screen. The Taylor expansion of the square root in equation (1.154) is derived as

$$\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \cdots \approx 1 + \frac{x}{2}$$

We retain up to second order; thus,

$$s \approx z + \frac{1}{2z}\left[(x-\xi)^2 + (y-\eta)^2\right] \tag{1.155}$$

© 2015 by Taylor & Francis Group, LLC

As a result of the paraxial approximation, an observation point should be close to the optical axis z. The first order development of s is valid if $x^2+y^2 \ll z^2$ and $\xi^2+\eta^2 \ll z^2$, which are not very stringent conditions. This approximation is equivalent to changing the emitted spherical wave into a quadratic wave. The diffracted field is written as

$$U(x,y) = \frac{e^{i\kappa z}}{i\lambda z} \iint_{-\infty}^{\infty} U(\xi,\eta) e^{i(\kappa/2z)\left[(x-\xi)^2+(y-\eta)^2\right]} d\xi d\eta \tag{1.156}$$

Equation (1.156) is seen to be a convolution equation (see Section 8.2.3), expressible in the form of

$$\begin{aligned} U(x,y) &= \iint_{-\infty}^{\infty} U(\xi,\eta)\, U_z(x-\xi, y-\eta)\, d\xi d\eta \\ &= U(\xi,\eta) \star U_z(x,y) \end{aligned} \tag{1.157}$$

in which $\star$ denotes the convolution parameter and

$$U_z(x,y) = \frac{e^{i\kappa z}}{i\lambda z} e^{i(\kappa/2z)\left[x^2+y^2\right]} \tag{1.158}$$

In U_z, two phase factors appear[146]:

- the first one corresponds to the general phase retardation as the wave travels from one plane to the other and
- the second one is a quadratic phase term that depends on the positions `O` and `P`.

When a monochromatic point source is obscured by a straight edge, for example, the edge of the Moon, the expected intensity pattern is described in terms of the Fresnel integrals:

$$\mathcal{C}(\mathrm{w}) = \int_0^{\mathrm{w}} \cos\left(\frac{\pi}{2}\tau^2\right) d\tau; \qquad \mathcal{S}(\mathrm{w}) = \int_0^{\mathrm{w}} \sin\left(\frac{\pi}{2}\tau^2\right) d\tau \tag{1.159}$$

in which $\mathrm{w} = x(2/\lambda L)^{5/2}$ is the Fresnel number (dimensionless), λ the wavelength of light, $L(= 384,000$ km) the distance from the Moon to Earth, and x the distance in meters from the observer to the edge of the lunar shadow.

As the Moon moves across the source, the fringe pattern moves across the telescope aperture as well and a light curve is observed. The irradiance, I, is given by[105]

$$I(\mathrm{w}) = \frac{I(0)}{2}\left[\left(\frac{1}{2}+\mathcal{C}(\mathrm{w})\right)^2 + \left(\frac{1}{2}+\mathcal{S}(\mathrm{w})\right)^2\right] \tag{1.160}$$

in which $I(0)$ is the unobstructed irradiance.

© 2015 by Taylor & Francis Group, LLC

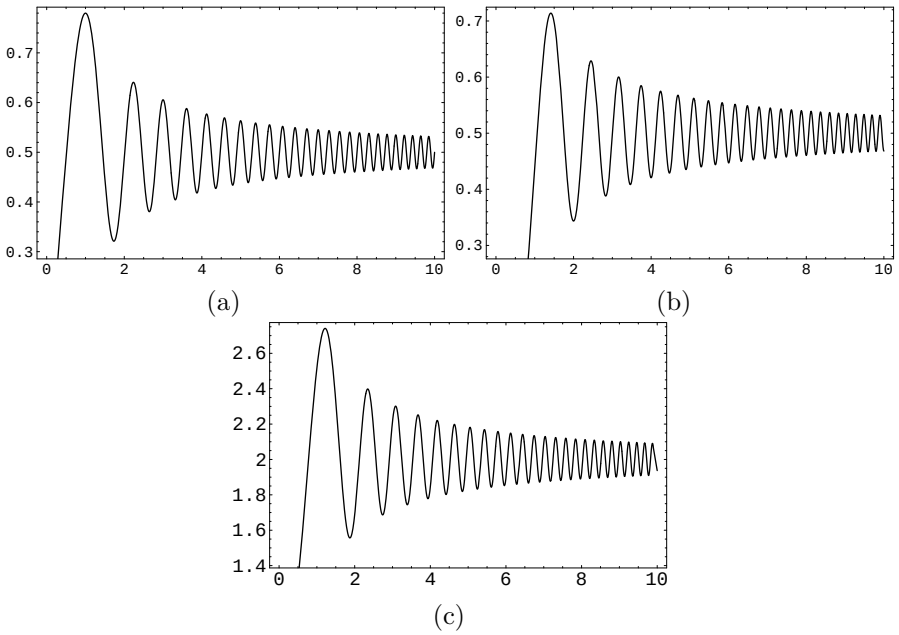


FIGURE 1.19: (a) Fresnel diffraction curves for $\mathcal{C}(\mathrm{w})$, (b) $\mathcal{S}(\mathrm{w})$, and (c) the monochromatic Fresnel diffraction patterns.

Figure 1.19 depicts the curves generated for (a) $\mathcal{C}(\mathrm{w})$ and (b) $\mathcal{S}(\mathrm{w})$, and (c) the monochromatic Fresnel diffraction patterns using equation (1.160). This light-curve embeds information about the 1D brightness profile of the source along the direction perpendicular to the lunar limb.

1.7.6 Fraunhofer Approximation

The Fraunhofer (or far-field) approximation, named after Joseph von Fraunhofer (1787−1826), takes place if the observation field (screen) is located further away from the diffraction screen. In this approximation, the factor $[\cos(\hat{\mathbf{n}}, \mathbf{s}) - \cos(\hat{\mathbf{n}}, \mathbf{r})]$ in equation (1.146) does not vary appreciably over the aperture. The diffraction pattern changes uniformly in size as the viewing screen is moved relative to the aperture. If $z \to \infty$, we can further simplify equation (1.155) as

$$s \approx z + \frac{1}{2z}\left[(x^2 + y^2\right] - \frac{1}{z}\left[x\xi + y\eta\right] \tag{1.161}$$

This is called the infinite distance approximation leading to the Fraunhofer diffraction. The conditions of validity are $x^2 + y^2 \ll z^2$ and $\xi^2 + \eta^2 \ll 2z/\kappa = \lambda z/\pi$, which are restrictive. The distance separating the Fresnel and the Fraunhofer regions is called the Rayleigh distance, $z_R = D^2/\lambda$, in which D is the

© 2015 by Taylor & Francis Group, LLC

size of the diffracting aperture. In the Fraunhofer diffraction case, we get

$$U(x,y) = \frac{e^{i\kappa z}}{i\lambda z} e^{i(\kappa/2z)\left[x^2+y^2\right]} \iint_{-\infty}^{\infty} U(\xi,\eta)\, e^{-i(\kappa/z)[x\xi+y\eta]} d\xi d\eta \tag{1.162}$$

Equation (1.162) does not include the nonlinear phase variation on the diffraction plane as it occurs in the Fresnel pattern; the observing plane is far away from the diffraction pattern, i.e., $z \gg (\xi^2+\eta^2)/\lambda$. In practice, to observe the Fraunhofer diffraction pattern by an aperture, one may use a lens for focusing the diffraction pattern onto a plane where the light distribution is the Fourier transform (FT; see Section 1.3.3) of the aperture. Retaining $(\xi^2+\eta^2)$ in equation (1.155), that is,

$$(x-\xi)^2 + (y-\eta)^2 = \left(x^2+y^2\right) - 2\,(x\xi + y\eta) + \xi^2 + \eta^2$$

Equation (1.156) may be written as

$$U(x,y,z) \propto \iint_{-\infty}^{\infty} U(\xi,\eta)\, e^{i(\kappa/2z)\left[\xi^2+\eta^2\right]} e^{-i(\kappa/z)[x\xi+y\eta]} d\xi d\eta \tag{1.163}$$

The Fraunhofer diffraction pattern for the field is proportional to the FT of the transmission function. It is, in general, complex since both the amplitude and the phase of the light may be altered on passing through the object. By introducing new coordinates

$$u = \frac{x}{\lambda z}; \qquad v = \frac{y}{\lambda z} \tag{1.164}$$

the Fraunhofer diffraction integral (equation 1.162) may be expressed in the form

$$\widehat{U}(u,v) = \mathcal{C} \iint_{-\infty}^{\infty} U(\xi,\eta)\, e^{-i2\pi(u\xi+v\eta)} d\xi d\eta \tag{1.165}$$

with $\mathcal{C}\left[= \frac{e^{i\kappa z}}{i\lambda z} e^{i(\kappa/2z)\left[x^2+y^2\right]}\right]$ as constant.

The new coordinates u, v are called the spatial frequencies and have the units of inverse distance. In the time domain, this is analogous to the spectra in which frequency is the inverse of time. These spatial frequencies are usually two-dimensional (2D) for imaging applications.

1.7.6.1 Fraunhofer Diffraction by Square Aperture

Let us consider that O is the origin of a coordinate system at the center of a square/rectangular aperture of sides $2a$ and $2b$, and $O\xi$ and $O\eta$ are the axes parallel to the sides. If the aperture is illuminated by a unit amplitude, in a general incident monochromatic plane wave, the field distribution across

© 2015 by Taylor & Francis Group, LLC

the aperture is equal to the amplitude transmission function $t_A(\xi,\eta)$ which is given by

$$t_A(\xi,\eta) = \begin{cases} 1 & -a \mid \xi \mid \langle a; \quad -b \mid \eta \mid \rangle b \\ 0 & \text{otherwise} \end{cases} \tag{1.166}$$

The amplitude transmittance function of an aperture is defined as the ratio of the transmitted field amplitude $U_t(\xi,\eta,0)$ to the incident field amplitude $U_i(\xi,\eta,0)$ at each (ξ,η) in the $z=0$ plane. For a rectangular aperture, the amplitude transmitted is given by

$$t_A(\xi,\eta) = rect\left(\frac{\xi}{2a}\right) rect\left(\frac{\eta}{2b}\right) \tag{1.167}$$

The Fourier transform of the complex disturbance, $U(\xi,\eta)$ is evaluated according to equation (1.165)

$$\begin{aligned} \widehat{U}(u,v) &= \mathcal{C}\int_{-a}^{a}\int_{-b}^{b} e^{-i2\pi[u\xi+v\eta]}d\xi d\eta \\ &= \mathcal{C}\int_{-a}^{a} e^{-i2\pi u\xi}d\xi \int_{-b}^{b} e^{-i2\pi v\eta}d\eta \\ &= \mathcal{C}\widehat{U}(u)\widehat{U}(v) \end{aligned} \tag{1.168}$$

where the input wave is

$$U(\xi,\eta) = rect\left(\frac{\xi}{2a}\right) rect\left(\frac{\eta}{2b}\right) \tag{1.169}$$

and

$$\widehat{U}(u) = \int_{-a}^{a} e^{-i2\pi u\xi}d\xi \tag{1.170}$$

$$\widehat{U}(v) = \int_{-b}^{b} e^{-i2\pi v\eta}d\eta \tag{1.171}$$

On integrating these equations (1.170 and 1.171), we find

$$\begin{aligned} \widehat{U}(u) &= 2a\frac{\sin 2\pi ua}{2\pi ua} \\ \widehat{U}(v) &= 2b\frac{\sin 2\pi vb}{2\pi vb} \end{aligned} \tag{1.172}$$

The diffraction pattern related to the field distribution of a rectangular aperture is given by the Fourier transform of a rectangular distribution. This varies like the so-called the sinc function (see Figure 1.20a; solid curve), also called the sampling function, a function that arises in signal processing and the theory of Fourier transform

$$sinc(x) = \begin{cases} 1, & \text{for } x = 0 \\ \sin(\pi x)/(\pi x) & \text{otherwise} \end{cases} \tag{1.173}$$

© 2015 by Taylor & Francis Group, LLC

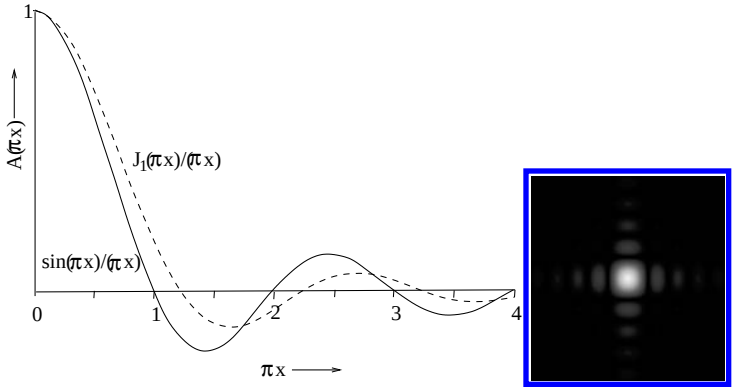


FIGURE 1.20: (a) Sinc function $sinc(x) = \sin(\pi x)/(\pi x)$ (solid line) and the function $J_1(\pi x)/(\pi x)$ (dashed line) and (b) 2D pattern of the Fraunhofer diffraction pattern at a square aperture.

which has the normalization

$$\int_{-\infty}^{\infty} sinc(x)dx = \pi$$

Therefore, the irradiance (intensity) at the point $\mathtt{P}(\mathbf{r})$ is given by

$$\widehat{I}(u,v) = |\widehat{U}(u,v)|^2 = \widehat{I}(0,0)sinc^2(2\pi ua)sinc^2(2\pi ub) \tag{1.174}$$

where $I(0,0) = E\mathtt{A}/\lambda^2$ is the intensity at the center of the pattern, E the total energy incident upon the aperture, and $\mathtt{A}(= 4ab)$ the area of the rectangle.

TABLE 1.1: Locations of the First Five Maxima of the Function[105]

x	$(\sin x/x)^2$
0	1
$1.430\pi = 4.493$	0.04718
$2.459\pi = 7.725$	0.01648
$3.470\pi = 10.90$	0.00834
$4.479\pi = 14.07$	0.00503

As stated in Section 1.5, the intensity is the measurable attribute of an optical wave field. The intensity of a monochromatic wave at point $\mathtt{P}$ is $I(\mathbf{r}) =\mid U(\mathbf{r}) \mid^2$, while for a narrow-band nonmonochromatic wave field it is

© 2015 by Taylor & Francis Group, LLC

$I(\mathbf{r}) = \langle |\ U(\mathbf{r},t)\ | \rangle^2$, where the instantaneous intensity $I(\mathbf{r},t) = |\ U(\mathbf{r},t)\ |^2$. The diffraction pattern in the case of a square aperture consists of a central bright spot (see Figure 1.20b) and a number of equi-spaced bright spots along the x and y axes, but with decreasing intensity. The intensity distribution of a function, $y = (\sin x/x)^2$, has a principal maximum $y = 1$ at $x = 0$ and zero minimum at $x = \pm\pi, \pm 2\pi, \pm 3\pi, \cdots$. The minima separate the secondary maxima whose locations are given in Table 1.1.

1.7.6.2 Fraunhofer Diffraction by Slit

If b is very large, the rectangular aperture degenerates into a slit, so that there is little diffraction in the y direction. The irradiance distribution in the diffraction pattern of a slit aperture is given by

$$I = I_0 sinc^2(2\pi ua) \tag{1.175}$$

The diffraction pattern of a slit aperture consists of a central bright line parallel to the slit and parallel bright lines of decreasing intensity on both sides.

1.7.6.3 Fraunhofer Diffraction by Circular Aperture

Let ρ, θ be the polar coordinates of a point in a circular aperture of radius a. The pupil function represented by $P(\rho,\theta)$ is

$$P(\rho,\theta) = \begin{cases} 1 & 0 < \rho < a \\ 0 & \text{otherwise.} \end{cases} \tag{1.176}$$

The circular polar coordinates are expressed in terms of u and v as

$$\begin{aligned} \xi &= \rho\cos\theta \qquad \eta = \rho\sin\theta \\ u &= w\cos\phi \qquad v = w\sin\phi \end{aligned} \tag{1.177}$$

with $w(= \sqrt{u^2+v^2})$ as the radial distance of the point P from the origin in units of length^{-1} and the sine of the angle which the direction (u,v) makes with the central direction $u = v = 0$.

Assuming circular symmetry for the optical field behind the aperture, $U(\rho,\theta) = \mathcal{C}P(\rho)$, the Fraunhofer diffraction integral takes the form

$$\widehat{U}(u,v) = \mathcal{C}\int_{\mathtt{W}} e^{-i2\pi(u\xi+v\eta)} d\xi d\eta \tag{1.178}$$

Using w in equation (1.178), we have

$$\begin{aligned} \widehat{U}(w) &= \mathcal{C}\int_0^a \rho d\rho \int_0^{2\pi} e^{-i2\pi\rho w\cos(\theta-\phi)} d\theta \\ &= 2\pi\mathcal{C}\int_0^a \rho J_0(-2\pi\rho w) d\rho \end{aligned} \tag{1.179}$$

© 2015 by Taylor & Francis Group, LLC

with $J_0(x)$ as the Bessel function of the first kind (see Section 1.7.6.6) and order zero and (w, ϕ) the coordinates of a point $\mathtt{P}(\mathbf{r})$ in the diffraction pattern with origin at the geometrical image of the source.

The disturbance of a circular aperture is expressed as

$$\widehat{U}(w) = \mathcal{C}\pi a^2 \left[\frac{2J_1(2\pi aw)}{2\pi aw}\right] \tag{1.180}$$

where $J_1(x)$ is the Bessel function of the first kind and order one.

The dashed curve of Figure 1.20a depicts the function $J_1(\pi x)/(\pi x)$.

1.7.6.4 Point Spread Function

The diffraction pattern is circularly symmetric; the intensity distribution is given by

$$\widehat{I}(w) = \widehat{I}(0)\left[\frac{2J_1(2\pi xw)}{2\pi xw}\right]^2, \tag{1.181}$$

where $\widehat{I}(0)[= \mathcal{C}^2(\pi x^2)^2]$ is the intensity on-axis ($w = 0$).

When normalized, the intensity distribution in the diffraction pattern through a circular aperture of a point source is a series of concentric rings with a bright central spot (see Figure 1.21). This intensity distribution is referred to as the Airy pattern, named after George B. Airy, which has an intense central maximum and a number of secondary maxima of considerably lower intensity. Table (1.2) shows the values of the Airy pattern at successive

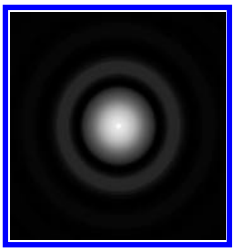


FIGURE 1.21: 2D pattern of the intensity distribution at a circular aperture.

maxima and minima.

The intensity distribution in diffraction pattern for pupils with or without aberrations is called the point spread function (PSF). It is defined as the illumination normalized by the illumination for $\theta = 0$ and is described by the modulus square of the Fourier transform of the aperture function. Thus, the resolution at the image plane is determined by the width of such PSF. Taking the property of the first order Bessel function, that is,

$$\lim_{x=0}\left[\frac{J_1(x)}{x}\right] = \frac{1}{2}$$

© 2015 by Taylor & Francis Group, LLC

TABLE 1.2: Locations of the Maxima and Minima of the Airy Pattern[105]

x	y $= \left[\frac{2J_1(x)}{x}\right]^2$	
0	1	Max.
$1.220\pi = 3.833$	0	Min.
$1.635\pi = 5.136$	0.0175	Max.
$2.233\pi = 7.016$	0	Min.
$2.679\pi = 8.417$	0.0042	Max.
$3.238\pi = 10.174$	0	Min.
$3.699\pi = 11.620$	0.0016	Max.

the PSF for a circular aperture can be recast as

$$\widehat{S}(w) = \left[\frac{2J_1(2\pi aw)}{2\pi aw}\right]^2 \tag{1.182}$$

In a circular beam, the spot has a sharp edge; hence, it is possible to specify spot diameter. Gaussian intensity profiles do not have sharp cutoffs at the edge that is specified as the point where the intensity reaches 13.5% of the maximum. The Fraunhofer diffraction pattern of a circular aperture has an easily defined edge. Such an edge of the circular spot where the Bessel function reaches its first zero is known as the first dark ring. This dark ring is called a region of destructive interference. For an imaging system, the PSF is proportional to the modulus square of the Fourier transform of its pupil transmission function; thus, the resolution at the image plane is determined by the width of this PSF.

1.7.6.5 Resolving Power of a Telescope

The Fraunhofer diffraction produced by the optical instruments that use circular pupils or apertures, say, telescope lenses or mirrors, plays an important role in the performances of these instruments. It is important to note that a telescope is an ideal condition for the image. When a continuum of wave components passes through telescope lenses or mirrors, the superposition of these components results in a pattern of constructive and destructive interference. For astronomical instruments, the incoming light is approximately a plane wave. A distant star produces spherical wavefronts and travels at the same speed in all directions. After traveling a long time, the radius of the wavefronts become so large that the wavefronts are considered to be planar over the aperture of a telescope. In tthe far-field limit, the Fraunhofer diffrac-

© 2015 by Taylor & Francis Group, LLC

tion occurs and the pattern formed at the focal plane of a telescope may have little resemblance to that at the aperture.

Let an atom be emitted at the surface of a star where the photon originated as an oscillator, a wavefront from which can be defined the collection of all points on the outwardly propagating wave, which are generated at the same time. An infinite series of concentric wavefronts is being emitted over time, any one of which can be chosen as the wave moves through the optical system. As the wavefront crosses the plane of the entrance pupil, the wavelets in the center of the pupil continue, propagating a plane wavefront. But at the edge of this pupil, there is no longer a reinforcement of wavelets from the part of the incident wavefront, and therefore, the wavelets inside the pupil commence spreading transversely into the geometric optics shadow region. In astronomical telescopes, one tries to find the wavelets in the far-field diffraction pattern. Another measurable quantity is, of course, the state of polarization of the electric field (see Section 8.4) of the detected photons.

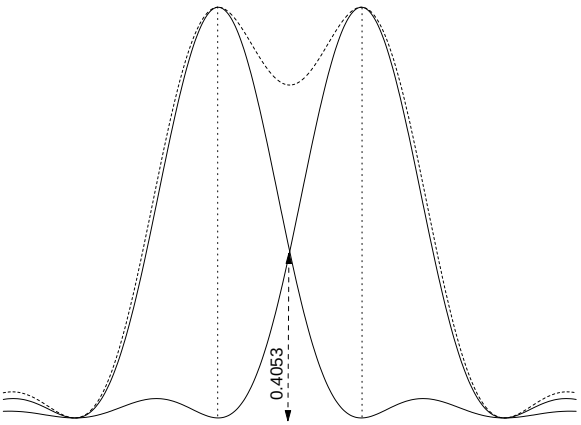


FIGURE 1.22: An example of imaging two point sources at the Rayleigh limit of separation.

The dimensions of the diffraction pattern are inversely proportional to the diameter of the entrance pupil. The deviation of the light does not occur, if the diameter of the aperture increases. But with the increase of the incident angle of the incoming beam, the entrance pupil may become occulted partially. This is known as the vignetting effect which modifies the distribution of light in the diffraction pattern. In addition, a partial occultation of the entrance pupil may occur due to the existence of a secondary mirror, thus producing a deformation and an enlargement of the diffraction pattern. In the absence of any other effects, such as, the effect of the atmospheric turbulence ([160] and references therein), the resolution of a telescope is limited by the diffraction of light waves. This is known as the diffraction-limit of a telescope, which is generally described by the following angle (in radians). Also known as the

© 2015 by Taylor & Francis Group, LLC

Airy disk, it is defined by

$$\theta_s \sim 1.22\frac{\lambda}{D} \tag{1.183}$$

where λ is a wavelength of light and D the diameter of the telescope.

The angular resolution, $\mathcal{R}$, of an imaging experiment, say, a telescope, is proportional to the wavelength of the radiation divided by the its diameter. The resolving power of any instrument, such as a telescope or an interferometer, refers to its ability to form images of two closely separated stars such that they are distinguishable. The smaller the diffraction disc, the higher the resolving power. Lord Rayleigh introduced a criterion, known as the Rayleigh criterion, for the resolving power of optical telescopes, which states that the two point sources may be resolved when the center maximum of the diffraction pattern of one source falls on the first minimum of the diffraction pattern of the other. At this separation, the separate contributions of the two sources to the intensity at the central minima is given by $4/\pi^2 = 0.4053$ for a rectangular aperture (see Figure 1.22).

In an ideal condition, the resolution that can be achieved in an imaging experiment, $\mathcal{R}$, is limited only by the imperfections in the optical system. For the measurement of such an imperfection,the Strehl's intensity ratio, that is, the ratio of the current peak intensity of a wave electric field to the peak intensity in the diffraction-limited seeing case where there are no phase fluctuations, was introduced by Karl Strehl. It is expressed as

$$\mathcal{S}_r = 1 - \left(\frac{2\pi}{\lambda}\right)^2 (\Delta\psi)^2_{RMS} \tag{1.184}$$

in which $\Delta\psi$ is the phase shift, and RMS stands for root-mean-square.

According to the Strehl's criterion, the resolving power, $\mathcal{R}$, of any telescope is given by the integral of its transfer function,

$$\mathcal{R} = \int_{-\infty}^{\infty} \widehat{S}(\mathbf{u}) d\mathbf{u} \tag{1.185}$$

Therefore, $\mathcal{R} = S(\mathbf{0})$. The Strehl's ratio, $\mathcal{S}_r$, is defined as the ratio of the intensity at the centroid of the observed PSF, $S(\mathbf{0})$, to the intensity of the peak of the diffraction-limited image or Airy spike, $S(\mathbf{0})_{\mathcal{A}s}$, that is,

$$\mathcal{S}_r = \frac{S(\mathbf{0})}{S(\mathbf{0})_{\mathcal{A}s}} \approx e^{-\kappa^2\sigma^2_{OPD}} \tag{1.186}$$

where κ is the wave number and σ_{OPD} the RMS optical path difference (OPD) error.

1.7.6.6 Bessel Function

The Bessel function of the first kind, $J_n(x)$, is the solution to the differential equation[100], which is nonsingular at the origin

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - n^2)y = 0 \tag{1.187}$$

© 2015 by Taylor & Francis Group, LLC

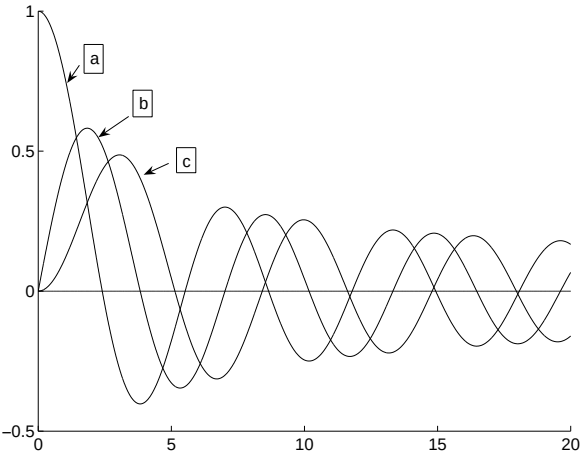


FIGURE 1.23: Bessel function of the first kind: (a) zero order, $J_0(x)$; (b) first order, $J_1(x)$; and (c) second order, $J_2(x)$.

Figure 1.23 depicts (a) the zero order, $J_0(x)$; (b) the first order, $J_1(x)$; and (c) the second order, $J_2(x)$ Bessel function of the first kind. The Bessel of the second kind, also called the Neumann function, $Y_n(x)$, is a solution to the differential equation (1.187) which is singular at the origin[129]. The Bessel function of the third kind, referred to as the Hankel function, is a linear combination of the Bessel functions of first and second kinds.

The boundary conditions of the Bessel function have circular symmetry:

$$J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} e^{ix\cos y}dy \tag{1.188}$$

$$J_n(x) = \frac{i^{-n}}{2\pi}\int_0^{2\pi} e^{i(ny+x\cos y)}dy \tag{1.189}$$

where n is the order (an integer), x the argument of the Bessel function, and

$$\frac{d}{dx}\left[x^n J_n(x)\right] = x^n J_{n-1}(x) \tag{1.190}$$

If $n = 1$, the integral identity emerges as

$$xJ_1(x) = \int_0^{2x} x' J_0(x')dx' \tag{1.191}$$

1.7.7 Fraunhofer Diffraction by Grating

Diffraction gratings are commonly used to isolate spectral regions in multi-channel instruments. These are often analyzed and synthesized by employing

© 2015 by Taylor & Francis Group, LLC

Fresnel and Fraunhofer approximations. A diffraction grating is a reflecting or transparent device, which periodically modulates the amplitude and the phase of an incident wave; both the amplitude and phase can be modulated as well. This modulation can be analyzed by means of the Fourier transform (see Section 1.3.3) into a number of sinusoidal components with certain amplitude coefficients. If the variation of transmitting power is sinusoidal, all the light of a plane wave incident on the grating goes into the two first-order spectra. If the n^{th} harmonic is present in the Fourier spectrum of the grating, there would be an n^{th}-order diffracted wave with amplitude proportional to the n^{th} Fourier coefficient.

A typical grating has fine parallel and equally spaced grooves or rulings, typically on the order of several hundreds per millimeter and has higher dispersion or ability to spread the spectrum than a prism does. The ruled gratings may be plano or concave and possess grooves, each parallel with the next. Holographic grating grooves can be either parallel or of unequal distribution in order that system performance may be optimized. Such gratings are generated on planar, spherical, toroidal, and many other surfaces. If a light beam strikes a grating, diffraction and mutual interference effects occur and light is reflected or transmitted in discrete directions, known as diffraction orders. Because of the dispersive properties of gratings, they are employed in monochromators and spectrometers. A grating creates monochromatic light from a white light source, which can be achieved by utilizing the grating's ability of spreading light of different wavelengths into different angles. The peak intensity occurs at diffraction angles governed by the following expression:

$$d \sin \theta_m = -m\lambda \tag{1.192}$$

in which m is an integer called the diffraction order, which may be positive or negative, resulting in diffracted orders on both sides of the zero order beam; θ_m the diffracted angle; d the groove spacing of the grating; $\lambda (= \lambda_0/n)$ the wavelength; λ_0 the wavelength in vacuum; and n the refractive index.

Equation (1.192) shows that at different wavelengths the beams travel in different directions. The resolving power, $R (= \lambda/d\lambda)$, in which $d\lambda$ is the difference in wavelength between two spectral lines of equal intensity, is a dimensionless quantity. Hence, the resolution is the ability of the instrument to separate adjacent spectral lines. In a blazed grating, the grooves of the diffraction grating are controlled to form right triangles with a blaze angle, the concentration of a limited region of the spectrum into any order other than the zero order. These gratings provide maximum efficiency at designated wavelengths.

The peak intensity at each wavelength depends on the number of slits in the grating. The transmission function of a grating is defined by

$$t(\xi, \eta) = \frac{U(\xi, \eta, 0_+)}{U(\xi, \eta, 0_-)} \tag{1.193}$$

© 2015 by Taylor & Francis Group, LLC

in which $U(\xi, \eta, 0_-)$ and $U(\xi, \eta, 0_+)$ are the respective wave functions immediately before and after the grating.

In the case of reflection grating, such a function turns out to be the reflection function. $t(\xi, \eta)$ represents a complex disturbance, at the pupil plane due to the monochromatic point source of frequency ν, of unit strength and of zero phase. If $arg \mid t(\xi, \eta) \mid = 0$, the grating modifies the amplitude, while in the case of $\mid t(\xi, \eta) \mid =$ constant, it modifies the phase.

1.7.7.1 Diffraction by Sinusoidal Amplitude Grating

An array of alternating opaque and transparent regions is called the transmission amplitude grating. Let us assume a plane wave incident perpendicularly on the grating with unit amplitude, the amplitude transmittance function is given by[119]

$$t_A(\xi, \eta) = \left[\frac{1}{2} + \frac{m}{2}\cos(2\pi u_0 \xi)\right] U(\xi, \eta) \tag{1.194}$$

where the grating structure is assumed to be bounded by a square aperture of width $2a$, the parameter m represents the peak-to-peak change of amplitude transmittance across the screen, u_0 is the spatial frequency of the grating, and $U(\xi, \eta)$ is defined by equation (1.169).

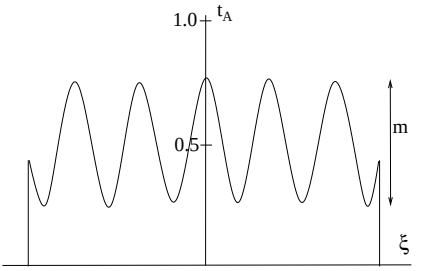


FIGURE 1.24: Amplitude transmittance function of the sinusoidal amplitude grating.

The term *thin* in this context means that the structure can be represented by a simple amplitude transmittance. Figure 1.24 depicts a cross-section of the grating amplitude transmittance function. The Fourier transforms of the term within square bracket on the RHS of equation (1.194), as well as $U(\xi, \eta)$, are respectively given by

$$\begin{aligned} \mathcal{F}\left\{\frac{1}{2} + \frac{m}{2}\cos(2\pi u_0 \xi)\right\} &= \frac{1}{2}\delta(u, v) + \frac{m}{4}\delta(u + u_0, v) \\ &\quad + \frac{m}{4}\delta(u - u_0, v) \qquad (1.195) \\ \widehat{U}(u, v) &= \mathtt{A} sinc(2au) sinc(2av) \qquad (1.196) \end{aligned}$$

© 2015 by Taylor & Francis Group, LLC

in which A signifies the area of the aperture bounding the grating.

The convolution theorem can be used to write

$$\begin{aligned}\widehat{U}(u,v) &= \mathcal{F}\{t_A(\xi,\eta)\} = \frac{\mathtt{A}}{2} sinc(2av)\{sinc(2au) \\ &\quad + \frac{m}{2} sinc[2a(u+u_0)] + \frac{m}{2} sinc[2a(u-u_0)]\Big\} \end{aligned} \tag{1.197}$$

The Fraunhofer diffraction pattern is expressed as

$$\begin{aligned} U(x,y) &= \frac{\mathtt{A}}{i2\lambda z} e^{i\kappa z} e^{(i\kappa/(2z))(x^2+y^2)} sinc\left(\frac{2b}{\lambda z}\right)\left\{sinc\left(\frac{2a}{\lambda z}\right)\right. \\ &\quad \left. + \frac{m}{2} sinc\left[\frac{2a}{\lambda z}(x+u_0\lambda z)\right] + \frac{m}{2} sinc\left[\frac{2a}{\lambda z}(x-u_0\lambda z)\right]\right\} \end{aligned} \tag{1.198}$$

The corresponding intensity distribution is derived by taking the square magnitude of equation (1.198). It is pertinent to note that if there are several gratings within the aperture, then $u_0 \gg 1/a$, and there would be insignificant overlap of the three sinc functions, allowing the intensity to be calculated as the sum of the squared magnitudes of the three terms. The intensity is given by

$$\begin{aligned} I(x,y) &\approx \left[\frac{\mathtt{A}}{2\lambda z}\right]^2 sinc^2\left(\frac{2b}{\lambda z}\right)\left\{sinc^2\left(\frac{2a}{\lambda z}\right)\right. \\ &\quad \left. + \frac{m^2}{4} sinc^2\left[\frac{2a}{\lambda z}(x+u_0\lambda z)\right] + \frac{m^2}{4} sinc^2\left[\frac{2a}{\lambda z}(x-u_0\lambda z)\right]\right\} \end{aligned} \tag{1.199}$$

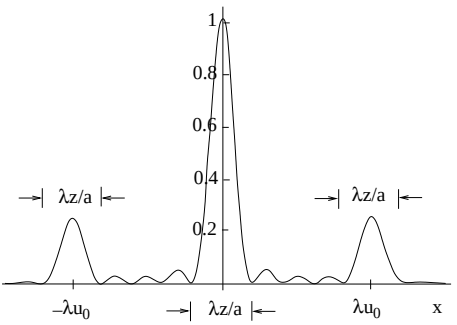


FIGURE 1.25: Fraunhofer diffraction pattern for a thin sinusoidal amplitude grating.

This intensity pattern is depicted in Figure 1.25. Some of the incident light

© 2015 by Taylor & Francis Group, LLC

may get absorbed by the grating, and also the sinusoidal transmittance variation across the aperture has deflected some of the energy out of the central diffraction pattern into two additional side patterns[127]. The central diffraction pattern is called the zero order of the Fraunhofer diffraction pattern, while the two side patterns are called the first orders. The spatial separation of the first orders from the zero order is $u_0 \lambda z$, while the width of the main lobe of all orders is $\lambda z/a$.

1.7.7.2 Diffraction by Sinusoidal Phase Grating

If the entire grating is transparent but varies periodically in optical thickness, it is referred to as a transmission phase grating. A reflective material with periodic surface relief can produce distinct phase relationships upon reflection of a wave. Such gratings are known as reflection phase gratings. With an incident plane wave on the grating, the transmission function for sinusoidal phase grating is given by

$$t_P(\xi, \eta) = e^{[i(m/2)\sin(2\pi u_0 \xi)]} U(\xi, \eta) \tag{1.200}$$

where by proper choice of phase reference, we have dropped a factor representing the average phase delay grating; $U(\xi, \eta)$ is defined by equation (1.169).

The analysis can be simplified by the following identity

$$e^{[i(m/2)\sin(2\pi u_0 \xi)]} = \sum_{n=-\infty}^{\infty} J_n\left(\frac{m}{2}\right) e^{i2\pi n u_0 \xi} \tag{1.201}$$

with $J_n(.)$ as the Bessel function of the first kind of order n.

The FT of equation (1.201) is derived as

$$\mathcal{F}\left\{e^{[i(m/2)\sin(2\pi u_0 \xi)]}\right\} = \sum_{n=-\infty}^{\infty} J_n\left(\frac{m}{2}\right) \delta(u - n u_0, v) \tag{1.202}$$

and therefore,

$$\begin{aligned} \widehat{U}(u,v) &= \mathcal{F}\{t_P(\xi,\eta)\} \\ &= \mathtt{A} sinc(2au) sinc(2av) \star \sum_{n=-\infty}^{\infty} J_n\left(\frac{m}{2}\right) \delta(u - n u_0, v) \\ &= \sum_{n=-\infty}^{\infty} \mathtt{A} J_n\left(\frac{m}{2}\right) sinc[2a(u - n u_0)] sinc(2av) \end{aligned} \tag{1.203}$$

where $\star$ denotes a convolution parameter.

Thus, the field strength in the Fraunhofer diffraction pattern is

$$\begin{aligned} U(x,y) &= \frac{\mathtt{A}}{i\lambda z} e^{i\kappa z} e^{i\kappa/(2z)(x^2+y^2)} \sum_{n=-\infty}^{\infty} J_n\left(\frac{m}{2}\right) \\ &\quad \times sinc\left[\frac{2a}{\lambda z}(x - n u_0 \lambda z)\right] sinc\left(\frac{2ay}{\lambda z}\right) \end{aligned} \tag{1.204}$$

© 2015 by Taylor & Francis Group, LLC

The corresponding intensity is

$$I(x,y) = \left(\frac{\mathtt{A}}{\lambda z}\right)^2 \sum_{n=-\infty}^{\infty} J_n^2\left(\frac{m}{2}\right) sinc^2\left[\frac{2a}{\lambda z}(x - nu_0\lambda z)\right] sinc^2\left(\frac{2ay}{\lambda z}\right) \tag{1.205}$$

Introduction of the sinusoidal phase grating has, therefore, deflected energy out of the zero order into a multitude of higher orders. The peak intensity of the n^{th} order is $[\mathtt{A}J_n(m/2)/\lambda z]^2$, while the displacement of that order from the center of the diffraction pattern is $nu_0\lambda z$. Figure 1.26 depicts a cross-section of

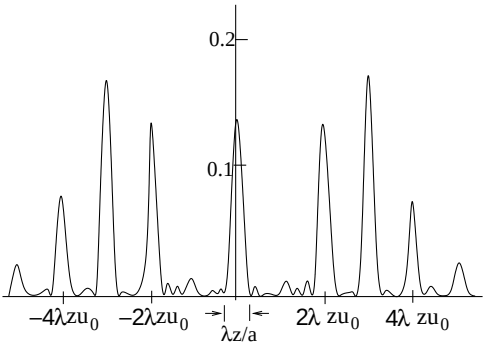


FIGURE 1.26: Fraunhofer diffraction pattern for a thin sinusoidal phase grating.

the intensity pattern when the peak-to-peak phase delay m is 8 radians[127]. The diffraction efficiency of n^{th} order of the grating is $\eta_n = J_n^2(m/2)$; the largest possible η into one of the $+1$ and -1 orders is the maximum value of J_1^2, which is 33.8%.

1.8 Exercises

1. Write short notes on
   (a) Point spread function (PSF).
   (b) Modulation transfer function (MTF).
   (c) Optical transfer function (OTF).
   (d) Strehl's ratio.
   (e) Dispersion of optical materials.
   (f) Diffraction limited of a telescope.
   (g) Coherence length.

© 2015 by Taylor & Francis Group, LLC

(h) Coherence time.

(i) Blazed reflection grating.

2. Determine the focal length and the diameter of the collimating lens in order to fill a grating of physical dimension of 30 mm in a grating spectrograph which is placed at an angle of 38.3° in an f/4 optical (telescope) system.

3. Light of wavelength 500 nm is incident on a diffraction grating of 1500 lines per centimeter. At what angle does the principal maximum occur?

4. What is the width of a slit if the first minimum for red light (650 nm) falls at $\theta = 30°$?

5. The equation of a plane perpendicular to the unit vector $\hat{\mathbf{s}}$ is $\hat{\mathbf{s}}\cdot\mathbf{r} =$ Constant. Find a plane wave solution of the three-dimensional wave equation given by equation (1.16).

6. Deduce the two-dimensional wave equation in polar coordinates (r, ϕ).

7. First order diffraction image of a slit is produced by a grating at 30° diffraction angle when light of 600 nm wavelength is incident on it. What is the number of lines per mm of the grating?

8. What will be the resolving power of a telescope objective of diameter 1 m for $\lambda = 589$ nm?

9. Refractive index (RI) of a specimen of glass is 1.5. A trough of this glass holds water (RI 1.33). What is critical angle between this glass and water?

10. Derive an expression for the intensity, I, of the superposition of two plane waves of wavelength, λ, traveling in opposite directions along the z axis.

11. On using an interference experiment with quasi-monochromatic approximation, express the visibility, $\mathcal{V}$, of the fringes as a function of the fields $U(\mathbf{r}_1, t)$ and $U(\mathbf{r}_2, t)$ of the sources. Also, show that the complex degree of coherence $|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| = \mathcal{V}$, in which $\mathcal{V}$ is the visibility.

12. Derive an expression for the visibility, $\mathcal{V}$ (see equation 1.112), as a function of the ratio $I(\mathbf{r}_1, t)/I(\mathbf{r}_1, t)$ for which the visibility is maximum.

13. Prove that when n is an integer,

(a) $J_{-n}(x) = (-1)^n J_n(x)$

(b) $x^2 \frac{d^2y}{dx^2} + x\frac{dy}{dx} + x^2 y = 0$

14. If $e^{x(z-1/z)/2} = \sum_{n=-\infty}^{\infty} J_n(x) z^n$, i.e., $e^{x(z-1/z)/2}$, is the generating function of $J_n(x)$, prove the following:

© 2015 by Taylor & Francis Group, LLC

(a) $xJ_n'(x) = xJ_{n-1}(x) - nJ_n(x)$

(b) $xJ_n'(x) = nJ_n(x) - xJ_{n+1}(x)$

(c) $\frac{d}{dx}\left[x^n J_n(x)\right] = x^n J_{n-1}(x)$

15. Prove the following Fourier transform (FT) relations:

(a) $\delta(a\xi, b\eta) = \frac{1}{|ab|}\delta(\xi, \eta)$

(b) $\mathcal{F}\{rect(x)rect(y)\} = sinc(u)sinc(v)$, in which u and v are the spatial frequencies and $\mathcal{F}$ the FT.

16. Prove and plot the FT of the following functions:

(a) $e^{-\pi x^2} \rightleftharpoons e^{-\pi u^2}$

(b) $sinc(x) \rightleftharpoons \Pi(u)$; $sinc\, x) = \dfrac{sinc\, \pi x}{\pi x}$
and $\Pi(x) = \begin{cases} 1 & \text{if } |x| < 1/2 \\ 0 & \text{if } |x| > 1/2 \end{cases}$

(c) $sinc^2(x) \rightleftharpoons \Lambda(u)$; $\Lambda(x) = \begin{cases} 1 - |x| & \text{if } |x| < 1 \\ 0 & \text{if } |x| < 1 \end{cases}$

(d) $\delta(x) \rightleftharpoons 1$

(e) $\text{III}(x) \rightleftharpoons \text{III}(u)$; $\text{III}(x) = \sum_{-\infty}^{\infty} \delta(x - n)$

17. Plot 2D Fourier transform of an 8×8 checker board.

18. Find an expression for the intensity distribution in the Fraunhofer diffraction pattern of a slit.

© 2015 by Taylor & Francis Group, LLC

Chapter 2

Electromagnetic Waves

2.1 Electrostatics 74
2.1.1 Point Charge: Coulomb's Law 74
2.1.2 Electric Energy of a System of Charges 75
2.1.3 Electric Fields 77
2.1.3.1 Field Distributions 77
2.1.3.2 Electric Dipole 78
2.1.3.3 Charge Distributions 79
2.1.4 Electric Flux and Gauss' Law 81
2.1.5 Electric Flux Density 83
2.1.6 Gauss' Law for Electrostatics 85
2.1.7 Potential Difference 90
2.1.7.1 Deriving Electric Field from Electric Potential . 91
2.1.7.2 Divergence of the Electric Field 93
2.1.7.3 Poisson and Laplace Equations 94
2.1.7.4 Curl of the Electric Field 95
2.1.8 Field from Line and Surface Charges 96
2.1.9 Current Density 98
2.2 Magnetostatics 99
2.2.1 Magnetic Flux 101
2.2.2 Ampère's Law 102
2.2.3 Biot–Savart Law 102
2.2.4 Magnetic Potential 104
2.2.4.1 Divergence and Curl of **B** 104
2.2.4.2 Magnetic Vector Potential 105
2.2.5 Lorentz Force 106
2.3 Time-Varying Fields 107
2.3.1 Faraday's Law 107
2.3.2 Inductance 108
2.3.2.1 Self-Inductance 108
2.3.2.2 Mutual Inductance 109
2.3.3 Resonant Circuits 110
2.4 Maxwell's Equations 114
2.4.1 Maxwell's First Law 114
2.4.2 Maxwell's Second Law 115
2.4.2.1 Conduction Current Density 115
2.4.2.2 Displacement Current 116
2.4.2.3 Ampère's Law with Maxwell Correction 117
2.4.3 Maxwell's Third Law 118
2.4.4 Maxwell's Fourth Law 119
2.4.5 Maxwell's Equations—Sinusoidal Fields 120
2.4.6 Continuity Equation of Charge 121
2.4.7 Boundary Conditions 122
2.5 Energy Flux of Electrodynamics 124
2.5.1 Poynting Vector 126
2.5.2 Energy Conservation Law of the Electromagnetic Field 127
2.6 Electromagnetic Field Equations 131

© 2015 by Taylor & Francis Group, LLC

2.6.1 General Electromagnetic Wave 133
2.6.2 Harmonic Time Dependence 136
2.6.3 Fourier Transform of Harmonic Equations 139
2.7 Antennas 140
2.7.1 Significance of Antenna Shape 141
2.7.2 Radiation of Electromagnetic Wave by Antenna 142
2.7.2.1 Electric and Magnetic Fields of Oscillating Hertzian Dipole 146
2.7.2.2 Radiation Fields of a Hertzian Dipole Antenna 148
2.7.2.3 Radiation Resistance of an Antenna 148
2.7.2.4 Quarter-Wave Monopole and Half-Wave Dipole 150
2.7.2.5 Power Gain of an Antenna 151
2.7.3 Antennas for Radio Astronomy 151
2.7.4 Waves through Ionosphere 153
2.8 Exercises 156

2.1 Electrostatics

Classical (nonquantum) electromagnetic theory, in which the quantum law with its constant, h, is ignored, deals with electric charges and currents and their interactions. The theory is based on an experiment that is secure within its range of application to coils, capacitors, oscillating currents, and eventually radio and optical waves. We can describe the repulsion and attraction of particles using the same laws that apply to the leaves of an electroscope; quantum mechanics is needed to predict the behavior of the particles under those forces. Without going into details of the experiments by which the existence of electric charge, $q(= 1.6 \times 10^{-19}$ coulomb; C) was demonstrated, we shall enumerate the basic laws of stationary electric charges, called the electrostatics; the source of the static electric field is a stationary charge. The fundamental property of this charge is its existence in two varieties, positive (+) and negative (−). The charged particles are created in pairs with equal and opposite charge. The total electric charge in an isolated system does not change.

The state of excitation due to electric charges constitutes an electromagnetic field that is specified in any physical quantity which takes on different values at different points in space. This electric charge is quantized, as well as conserved. The strength of a particle's electric interaction with an object around it depends on its electric charge. In what follows, we describe the basic laws governing static electric and magnetic fields.

2.1.1 Point Charge: Coulomb's Law

Based on the direct measurement of the electrostatic force between the charged bodies, Charles Augueste Coulomb (1736−1806) announced the inverse squared law of electrical attraction and repulsion after measuring the

© 2015 by Taylor & Francis Group, LLC

force between small charged spheres[39] with a torsion balance. The electrostatic force, $\mathbf{F}$ (newton, N), acts on a test charge, q_1, located at $\mathbf{r}_1$, due to another point charge, q_2, at position $\mathbf{r}_1$, which is equal to ([130],[133])

$$\begin{aligned}\mathbf{F} &\propto q_1 q_2 \frac{\mathbf{r}_1 - \mathbf{r}_2}{\mid \mathbf{r}_1 - \mathbf{r}_2 \mid^3} = k q_1 q_2 \frac{\hat{\mathbf{r}}_{21}}{\mid \mathbf{r}_{21} \mid^2} \\ &= \frac{q_1 q_2}{4\pi\epsilon} \frac{\hat{\mathbf{r}}_{21}}{\mid \mathbf{r}_{21} \mid^2} \end{aligned} \tag{2.1}$$

where $k(= 1/4\pi\epsilon)$ is the constant of proportionality, $\mathbf{r}_{21}(= \mathbf{r}_1 - \mathbf{r}_2)$ the separation vector, $\hat{\mathbf{r}}_{21}(= \mathbf{r}_{21}/|\mathbf{r}_{21}|)$ the unit vector, and ϵ the permittivity of the medium in $\mathrm{F.m}^{-1}$.

Equation (2.1) expresses the fact that like charges repel and unlike charges attract. The force is proportional to the product of the magnitude of these charges and inversely proportional to the square of their separation. It obeys the third law devised by Issac Newton (1642−1727), i.e., $\mathbf{F}_2 = -\mathbf{F}_1$ (see Table A.1 in Appendix A).

Let us have some charges, $q_1, q_2, q_3, \cdots, q_N$, at respective positions, $\mathbf{r}_1, \mathbf{r}_2, \cdots \mathbf{r}_N$, from a small charge q at position $\mathbf{r}$. The field at a point is the vector sum of the fields of each of the charges in the system. The total force on q is

$$\begin{aligned}\mathbf{F}(\mathbf{r}) &= \mathbf{F}_1(\mathbf{r}_1) + \mathbf{F}_2(\mathbf{r}_2) + \mathbf{F}_3(\mathbf{r}_3) + \cdots \\ &= \frac{1}{4\pi\epsilon_0}\left(\frac{q_1 q}{\mid \mathbf{r}_1 \mid^2}\hat{\mathbf{r}}_1 + \frac{q_2 q}{\mid \mathbf{r}_2 \mid^2}\hat{\mathbf{r}}_2 + \frac{q_3 q}{\mid \mathbf{r}_3 \mid^2}\hat{\mathbf{r}}_3 + \cdots\right) \\ &= \frac{q}{4\pi\epsilon_0}\left(\frac{q_1\hat{\mathbf{r}}_1}{\mid \mathbf{r}_1 \mid^2} + \frac{q_2\hat{\mathbf{r}}_2}{\mid \mathbf{r}_2 \mid^2} + \frac{q_3\hat{\mathbf{r}}_3}{\mid \mathbf{r}_3 \mid^2} + \cdots\right) \end{aligned} \tag{2.2}$$

Since the charges, $q_1, q_2, q_3, \cdots, q_N$, fixed in space, exert force on the charge, q, that might be brought into their vicinity, equation (2.2) is recast as

$$\mathbf{F}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0}\sum_{j=1}^{N}\frac{q_j}{\mid \mathbf{r}_j \mid^2}\hat{\mathbf{r}}_j \tag{2.3}$$

with $j = 1, 2, 3, \cdots, N$; $\mathbf{r}(= x, y, z)$ as the position vector, $\hat{\mathbf{r}}_j$ the unit vector in the direction $\mathbf{r}_j = \mathbf{r} - \mathbf{r}_j$, a vector pointing charges q_j to q, and ϵ_0 the permittivity or dielectric constant at vacuum.

2.1.2 Electric Energy of a System of Charges

Given the charges and their locations, we are able to determine the electrical forces. We know that the electrical forces are conservative; therefore, even when we push charges around in electric fields, no energy is irrecoverably lost. Everything is reversible. Let us consider that two charged bodies are carrying charges q_1 and q_2 with position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. The force that

© 2015 by Taylor & Francis Group, LLC

is to be applied to move one charge toward the other is equal to and opposite the Coulomb force. The work of a system of two point charges, W, is defined as the integral of the force multiplied by the distance, that is,

$$\begin{aligned} W_2 &= \frac{1}{4\pi\epsilon_0}\int_{r=\infty}^{r_{12}} \frac{q_1 q_2(-dr)}{r^2} \\ &= \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r_{12}} \end{aligned} \tag{2.4}$$

where r_{12} is the distance from point 1 to point 2.

Since r is changing from ∞ to r_{12}, the increment of displacement is $-dr$. If q_1 and q_2 have the same sign, positive work must be done to overcome the electrostatic repulsion and the change in the potential energy, w, of the system is positive. Let us consider a charge q_1 kept fixed but the charge q_2 is moved to a position P along two different paths crossing a spherical shell of which dr is its thickness. If r is the distance between the charge q_1 and the inner edge of the shell, both paths would cross between r and $r + dr$. The increment of work involved, $-\mathbf{F}\cdot d\mathbf{l}$ in this part of the path, is the same for the two paths. Here scalar product (dot product; see Appendix B.1.1) of two vectors is used beacuse $\mathbf{F}$ has the same magnitude at both places and is radially directed from q_1, while $d\mathbf{l} = dr/\cos\theta$; hence, $\mathbf{F}\cdot d\mathbf{l} = F dr$. Each increment of work along one path is matched by a corresponding increment on the other; thus, the sums must be equal. The work of a system of three point charges would be

$$W_3 = -\int_{\infty}^{\mathrm{P}_3} \mathbf{F}_3\cdot d\mathbf{l} \tag{2.5}$$

hence, we can write

$$\begin{aligned} -\int \mathbf{F}_3\cdot d\mathbf{l} &= -\int (\mathbf{F}_{31} + \mathbf{F}_{32})\cdot d\mathbf{l} \\ &= -\int \mathbf{F}_{31}\cdot d\mathbf{r} - \int \mathbf{F}_{32}\cdot d\mathbf{r} \end{aligned} \tag{2.6}$$

That is, the work required to bring q_3 to P_3 is the sum of the work needed when q_1 is present alone and that needed when q_2 is present alone:

$$W_3 = \frac{q_3}{4\pi\epsilon_0}\left(\frac{q_1}{r_{31}} + \frac{q_2}{r_{32}}\right) \tag{2.7}$$

The total work done, w, in assembling this arrangement of three charges is therefore

$$W = \frac{1}{4\pi\epsilon_0}\left(\frac{q_1 q_2}{r_{12}} + \frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}}\right) \tag{2.8}$$

The electrical potential energy of this system, w, which is the total work required to assemble the system, is independent of the order. This result can

© 2015 by Taylor & Francis Group, LLC

be generalized to any number of charges. For N different charges, the potential energy, w, is derived by summing over all pairs as in equation (2.7), that is,

$$w = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}} \tag{2.9}$$

in which the double-sum notation, $\sum_{i=1}^{N}\sum_{j\neq i}^{N}$, implies the following: while taking $i = 1$, do sum over $j = 2, 3, 4, \cdots, N$, which is followed by $i = 2$ and sum over $j = 1, 3, 4, \cdots, N$ and so on, through $i = N$. One may count each pair twice, for which $1/2$ is put in front of the factor.

2.1.3 Electric Fields

An electric field, $\mathbf{E}$, is defined as the electric force per unit charge, $\mathbf{F}/q$, the concept of which was introduced by Michael Faraday (1791−1867). It is a vector field, which is to charge as gravitational acceleration is to mass. This field can be generated by electric charge, also referred to as source charge, q, as well as by a time-varying magnetic field. The source charge can be a single charge or a group of discrete charges or any continuous distribution of charge. The direction of the field is taken to be the direction of the force it would exert on a positive test charge. The electric field is radially outward from a positive charge and radially in toward a negative point charge. The electric field, at the point $\mathbf{r}$ due to a point charge q at the point $\mathbf{r}_1$, can be obtained as

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0}\frac{\mathbf{r}-\mathbf{r}_1}{\mid \mathbf{r}-\mathbf{r}_1 \mid^3} \tag{2.10}$$

The unit of the electric field, $\mathbf{E}$, is commonly stated in volts per meter, $\mathrm{V.m^{-1}}$ or newtons per coulomb ($\mathrm{N.C^{-1}}$); in the international system of units (SI), the units of newtons per coulomb and volts per meter are equivalent. The observed linear superposition of force due to many charges implies that we can write the electric field, $\mathbf{E}$, at $\mathbf{r}$ due to a system of point charges q_j located at $\mathbf{r}_j$, as the vector sum

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\sum_{j=1}^{N} q_j\frac{\mathbf{r}-\mathbf{r}_j}{\mid \mathbf{r}-\mathbf{r}_j \mid^3} \tag{2.11}$$

2.1.3.1 Field Distributions

The field due to multiple charges may be found by summing up all the individual fields. We may represent electric fields by drawing field lines in the direction of the electric field. The strength of the field is represented by the spacing of the lines; the closer the spacing of the lines, the stronger the field. The force on a charge, q, in an electric field is $\mathbf{F} = q\mathbf{E}$. The electric fields can be used to accelerate electrons and control their paths. In an electron gun,

© 2015 by Taylor & Francis Group, LLC

electrons are generated by a thermal emission process and are accelerated by an electric field. An electron can be accelerated to a very high speed in a short distance. The acceleration, **a**, produced by the field would be

$$\mathbf{a} = \frac{\mathbf{F}}{m} = \frac{q\mathbf{E}}{m} \tag{2.12}$$

where m is the mass of an electron.

We can derive the velocity, v, using the kinematic equation of motion, $v^2 = v_0^2 + 2ax$, in which v_0 is the initial velocity and x the distance. Here, $v = \sqrt{2ax}$.

2.1.3.2 Electric Dipole

A pair of point charges with equal and opposite sign which are separated by some fixed distance, d, are called an electric dipole; the electric field intensity at a distance $r \gg d$ is needed. A configuration of practical importance separated a short distance apart is depicted in Figure 2.1a. The electric field of a dipole, $\mathbf{E}_{dipole}$, is given by

$$\mathbf{E}_{dipole} = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}}{r^3} \tag{2.13}$$

with $\mathbf{p}(= q\mathbf{d})$ as the dipole moment.

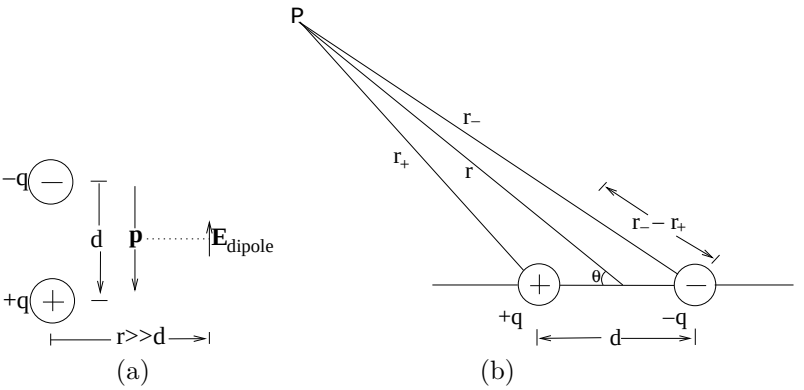


FIGURE 2.1: (a) The dipole configuration and (b) the electric dipole and calculation of potential at a point P due to the dipole.

The electric dipole moment for a pair of opposite charges of magnitude, q, is defined as the magnitude of the charge times the distance between them, and the defined direction is toward the positive charge. It is a useful concept in atoms and molecules where the effects of charge separation are measurable, but the distances between the charges are too small to be easily measured. It is also a useful concept in dielectrics and other applications in solid and liquid materials.

© 2015 by Taylor & Francis Group, LLC

Knowing the electrical potential function at that point as potential is of a scalar nature, one can derive the electric field intensity at a distant point. It is indeed easier to evaluate the electric potential than to evaluate the electric field intensity, which is a vector quantity.

Let us consider an electric dipole along the y axis, as depicted in Figure 2.1b. The electric potential, V, at a point P in the $x-y$ plane is given by

$$V = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{r_+} - \frac{1}{r_-}\right) = \frac{q}{4\pi\epsilon_0}\left(\frac{r_- - r_+}{r_+ r_-}\right) \tag{2.14}$$

The potential of a dipole is of most interest where $r \gg d$; hence, this can be approximated by

$$V = \frac{p\cos\theta}{4\pi\epsilon_0 r^2} = \frac{\mathbf{p}\cdot\hat{\mathbf{r}}}{4\pi\epsilon_0 r^2} \tag{2.15}$$

The standard approximations are $r_- - r_+ \approx d\cos\theta$ and $r_+ r_- \approx r^2$. A dipole in an electric field does not experience a net force; however, if the dipole is not aligned with the field, it experiences a torque, $\tau(= rF\sin\theta)$. The torque of each force about the center of the body is the product of the distance from the center, the magnitude of the force, and the sine of the angle between the line from the center and the force. Thus, the magnitude of the torque due to each force is $\tau = (a\,qE\sin\theta)$, and the net torque of both forces together is given by

$$\tau = -d\,qE\sin\theta$$

Here, the torque is clockwise; hence, the right-hand side (RHS) of this expression possesses a negative sign. The torque tends to align the body with the electric field. We can write this expression in the form

$$\tau = -pE\sin\theta \tag{2.16}$$

Expressed in coulomb-meters (C.m), this dipole moment p, is the magnitude of dipole moment vector, $\mathbf{p}$, whose direction is from the $-ve$ charge toward the $+ve$ charge. The torque vector is expressed in terms of the electric dipole moment $\mathbf{p}$ as

$$\boldsymbol{\tau} = \mathbf{p}\times\mathbf{E} \tag{2.17}$$

where the direction of $\boldsymbol{\tau}$ is the axis of rotation.

The potential energy is the negative of the work done, $w = -W$, so

$$w = -pE\cos\theta = -\mathbf{p}\cdot\mathbf{E} \tag{2.18}$$

2.1.3.3 Charge Distributions

Often the distances between charges in a group of charges are much smaller than the distance from the group to some point of interest. In such situations, the system of charges become continuous, which means the system of closely spaced charges is equivalent to a total charge, i.e., continuously distributed

© 2015 by Taylor & Francis Group, LLC

along some line, over some surface, or throughout some volume. Equation (2.11) assumes that the source of the field is a collection of discrete point charges q_i. If the charge is distributed continuously over some region, the sum becomes an integral

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\mathbf{r}-\mathbf{r}_1}{\mid \mathbf{r}-\mathbf{r}_1 \mid^3} dq \tag{2.19}$$

To note, the integral[1] provides the electric field at $\mathbf{r}$, which is produced at other points $\mathbf{r}'$. In order to find the electric field due to a continuous charge distribution, we have to define the following terms:

1. Linear charge density: If the charge is uniformly distributed along a line with charge per unit length $\lambda(\mathbf{r}')$ having unit of C.m^{-1}, then $dq = \lambda dl'$, in which dl' is an element of length along the line. The electric field of a line charge is

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int_{\mathtt{P}} \lambda(\mathbf{r}')\frac{\mathbf{r}-\mathbf{r}'}{\mid \mathbf{r}-\mathbf{r}' \mid^3} dl' \tag{2.20}$$

2. Surface charge density: If the charge is smeared out over a surface with charge per unit area $\sigma(\mathbf{r}')$ having unit of C.m^{-2}, then $dq = \sigma d\mathtt{S}'$, where $d\mathtt{S}$ is an element of surface along the area on the surface. The differential element of area on a spherical surface in spherical coordinates is $d\mathtt{S} = r^2 \sin\theta d\theta d\phi$, in which (θ, ϕ) are the polar coordinates. The electric field for a surface charge is

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int_{\mathtt{S}} \sigma(\mathbf{r}')\frac{\mathbf{r}-\mathbf{r}'}{\mid \mathbf{r}-\mathbf{r}' \mid^3} d\mathtt{S}' \tag{2.21}$$

3. Volume charge density: If the charge fills a volume with charge per unit volume $\rho(\mathbf{r}')$ having unit of C.m^{-3}, then $dq = \rho d\mathtt{V}'$, with $d\mathtt{V}'$ as an element of volume. The electric field for a volume charge is

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int_{\mathtt{V}} \rho(\mathbf{r}')\frac{\mathbf{r}-\mathbf{r}'}{\mid \mathbf{r}-\mathbf{r}' \mid^3} d\mathtt{V}' \tag{2.22}$$

   with $\mathtt{V}$ as the volume of the charged object.

[1] Integration over a distribution of charge, $\int dq$, can be performed in the following ways:

- for a line charge: $\int dq = \int \lambda dx$,
- for a sheet of charge: $\int dq = \int \sigma d\mathtt{S}$, and
- for volume of charge: $\int dq = \int \rho d\mathtt{V}$, in which λ, σ, and ρ are the charges per unit length, per unit area, and per unit volume, respectively.

Depending on the symmetry of the problem, we may choose to integrate over Cartesian coordinates or polar coordinates, for example,

- for a rectangular sheet of charge: $\int dq = \int \sigma d\mathtt{S} = \iint \sigma dx dy$ and
- for a circular sheet of charge: $\int dq = \int \sigma d\mathtt{S} = \iint \sigma r dr d\theta$.

© 2015 by Taylor & Francis Group, LLC

Here,

$$dq \to \lambda dl' \sim \sigma d\mathtt{S}' \sim \rho d\mathtt{V}'$$

2.1.4 Electric Flux and Gauss' Law

The term *flux* is a quantity expressing the strength of a field of force in a given area. The flux passing through any closed surface in an electric field is equal to the algebraic sum of the electric charges enclosed by that surface. This statement is referred to as Gauss' law for electrostatics, named after Karl Friedrich Gauss (1777−1855). This closed surface is referred to as a Gaussian surface. Consider a point charge q and a closed surface $\mathtt{S}$, as depicted in Figure 2.2. Let r be the distance from the charge to a point on the surface, $\hat{\mathbf{n}}$ be the outwardly directed unit normal vector to the surface at the point, and $d\mathtt{S}$ be an element of surface area within which the field is uniform. If the electric field, $\mathbf{E}$, at the point on the surface due to the charge, q, makes an angle, θ, with the unit normal, the normal component of $\mathbf{E}$ times the area element is[133]

$$\mathbf{E}\cdot\hat{\mathbf{n}}d\mathtt{S} = \frac{q}{4\pi\epsilon_0}\frac{\cos\theta}{r^2}d\mathtt{S} \tag{2.23}$$

$\mathbf{E}$ is directed along the line from the surface element to the charge q; thus, $\cos\theta d\mathtt{S} = r^2 d\Omega$, in which $d\Omega$ is the element of solid angle subtended by $d\mathtt{S}$ at the position of charge. Hence,

$$\mathbf{E}\cdot\hat{\mathbf{n}}d\mathtt{S} = \frac{q}{4\pi\epsilon_0}d\Omega. \tag{2.24}$$

Here,

$$\oint_{\mathtt{S}} \mathbf{E}\cdot\hat{\mathbf{n}}d\mathtt{S} = \begin{cases} q/\epsilon_0 & \text{if q is inside } \mathtt{S} \\ 0 & \text{otherwise} \end{cases} \tag{2.25}$$

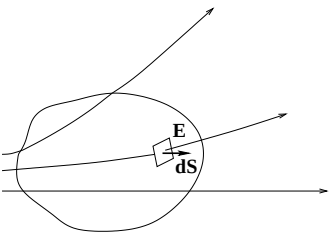


FIGURE 2.2: Electric flux for non-uniform fields.

For a discrete set of charges, we can divide the arbitrary surface and field into infinitesimally small areas, $d\mathtt{S}$. The electric flux, Φ_E, measured in coulombs, C, is an additive quantity. If we have a number of sources, such as

© 2015 by Taylor & Francis Group, LLC

$q_1, q_2, q_3, \cdots, q_N$, the fields of which, if each of them were present alone, would be $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \cdots, \mathbf{E}_N$, the flux, Φ_E, through the surface, $\mathtt{S}$, is expressed as

$$\Phi_E = \int_{\mathtt{S}} (\mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 + \cdots + \mathbf{E}_N) \cdot d\mathbf{S} \tag{2.26}$$

The definition of flux is valid for a closed surface, for instance, the surface of a sphere or surface of a cube, as well. The surface integral of the normal component of electric field, $\mathbf{E}$, over any closed surface, $\mathtt{S}$, is equal to $1/\epsilon_0$ times the net charge, $\sum_j q_j$, enclosed by the surface. In an electric field, the electric flux tubes of a static electric field originate and end on electric charges, while the tubes of magnetic flux are continuous. Thus, according to Gauss' law

$$\Phi_E = \oint_{\mathtt{S}} \mathbf{E} \cdot \hat{\mathbf{n}} d\mathtt{S} \tag{2.27}$$

$$= \frac{1}{\epsilon_0} \sum_j q_j = \frac{1}{\epsilon_0} \int_{\mathtt{V}} \rho(\mathbf{r}) d\mathtt{V} \tag{2.28}$$

with $q = \oint \rho d\mathtt{V}$; ρ as the charge density, $\mathtt{V}$ the volume, and $d\mathtt{V}$ the volume element.

Equation (2.28) is one of the basic equations of electrostatics. It depends on several factors:

- the inverse square law for the force between charges,
- the central nature of the force, and
- the linear superposition of the effects of different charges.

The potential energy, w, of a system of charges is calculated from the electric field, that is,

$$w = \frac{\epsilon_0}{2} \int_{\text{entire field}} \mathbf{E}^2 d\mathtt{V} \tag{2.29}$$

In the case of two charged particles, such as protons, following equation (2.29), the total energy in the electric field[155] is given by

$$w = \frac{\epsilon_0}{2} \int (\mathbf{E}_1 + \mathbf{E}_2)^2 d\mathtt{V}$$

$$= \frac{\epsilon_0}{2} \int \mathbf{E}_1^2 d\mathtt{V} + \frac{\epsilon_0}{2} \int \mathbf{E}_2^2 d\mathtt{V} + \epsilon_0 \int \mathbf{E}_1 \cdot \mathbf{E}_2 d\mathtt{V} \tag{2.30}$$

in which $\mathbf{E}_1$ is the electric field of one particle and $\mathbf{E}_2$ that of the other with $\mathbf{E}(= \mathbf{E}_1 + \mathbf{E}_2)$ the total field.

In equation (2.30), the value of the first integral is a property of one particle alone, called the electrical self energy of one proton; the same holds for the second integral. The intrinsic property of the proton depends on its size and structure. The third integral involves the distance between the charges.

© 2015 by Taylor & Francis Group, LLC

Let us consider that $\mathbf{E}$ is the field of some stationary distribution of electric charges. Let $\mathtt{P}_1$ and $\mathtt{P}_2$ denote two points anywhere in the field. The line integral of $\mathbf{E}$ between two points is given by $\int_{\mathtt{P}_1}^{\mathtt{P}_2} \mathbf{E}{\cdot}d\mathbf{l}$, which is taken along some path passing from $\mathtt{P}_1$ to $\mathtt{P}_2$. Any electrostatic field can be regarded as the sum of a number of point charge fields as given by equation (2.11). If the line integral from $\mathtt{P}_1$ to $\mathtt{P}_2$ is independent of the path for each of the point charge fields $\mathbf{E}_1, \mathbf{E}_2, \cdots$, the total field is $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \cdots$. The line integral $\int_{\mathtt{P}_1}^{\mathtt{P}_2} \mathbf{E}{\cdot}d\mathbf{l}$ for any electrostatic field $\mathbf{E}$ has the same value for all paths from $\mathtt{P}_1$ to $\mathtt{P}_2$, that is,

$$\int_{\mathtt{P}_1}^{\mathtt{P}_2} \mathbf{E}{\cdot}d\mathbf{l} = \int_{\mathtt{P}_1}^{\mathtt{P}_2} \mathbf{E}_1{\cdot}d\mathbf{l} + \int_{\mathtt{P}_1}^{\mathtt{P}_2} \mathbf{E}_2{\cdot}d\mathbf{l} + \cdots \tag{2.31}$$

in which the same path is used for all the integrations.

If the points $\mathtt{P}_1$ and $\mathtt{P}_2$ coincide, the paths are all closed curves, which leads to the corollary: the line integral $\int \mathbf{E}{\cdot}d\mathbf{l}$ around any closed path in an electrostatic field is zero.

2.1.5 Electric Flux Density

As discussed in Section 2.1.4, Gauss' law holds for Newtonian gravitational force fields with matter density replacing charge density. The electric flux density, $\mathbf{D}$, measured in $\mathrm{C.m}^{-2}$ is a vector field, whose direction at a point is the direction of the flux lines at that point, and the magnitude is the number of flux lines crossing a surface normal to the lines divided by the surface area[145]. The total flux passing through the closed surface is obtained by adding the differential contributions crossing each surface element, $d\mathtt{S}$,

$$\Phi_E = \int d\Phi_E = \oint_{\mathtt{S}} \mathbf{D}{\cdot}d\mathbf{S} \qquad \text{coulomb (C)} \tag{2.32}$$

The integral of the normal component of the electrical flux density over any closed (Gaussian) surface is equal to the net charge enclosed. Equation (2.32) implies that the electric flux through any surface equals the integral of the flux density over the surface. The electric flux density, $\mathbf{D}$, is proportional to the electric field. The proportionality constant, ϵ, depends on the medium being analyzed:

$$\mathbf{D} = \epsilon\mathbf{E} = \frac{1}{4\pi}\frac{q}{r^2}\hat{\mathbf{r}} \tag{2.33}$$

The permittivity, ϵ, is constant in a linear, isotropic, and homogeneous medium. In an inhomogeneous medium $\epsilon = \epsilon(\mathbf{r})$, while in a nonlinear medium $\epsilon = \epsilon(\mathbf{E})$. For free space, the electric flux density is

$$\mathbf{D} = \epsilon_0\mathbf{E} \tag{2.34}$$

where ϵ_0 is the permittivity in free space.

In an anisotropic medium that implies the permittivity depends on which

© 2015 by Taylor & Francis Group, LLC

direction the fields are in, for example, a crystalline material, **D** may not be parallel to **E**:

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_x & 0 & 0 \\ 0 & \epsilon_y & 0 \\ 0 & 0 & \epsilon_z \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{2.35}$$

The electric flux density due to a point charge is the same at all points on the surface of a sphere of radius r; the flux is independent of the size of the sphere. The flux density vector is directed radially outwards from the source if the charge, q, is positive (+ve) and radially inward if q is negative (−ve). The magnitude of the electric flux density should be same at all points at equal distances from the charge, q. Thus, the magnitude of the electric flux density at any point at a distance, r, from a point source should be equal to the flux from the charge divided by the total area of a sphere of radius r. Therefore, the electric flux density, **D**, at a point in a nonconducting medium is

$$\mathbf{D} = \frac{\Phi_E}{4\pi r^2}\hat{\mathbf{r}} = \frac{q}{4\pi r^2}\hat{\mathbf{r}} \tag{2.36}$$

The net electric flux density or electric field strength at any point is the vector sum of the contributions from each point charge. From the principle of superposition, we get

$$\mathbf{D} = \mathbf{D}_1 + \mathbf{D}_2 + \mathbf{D}_3 + \cdots \tag{2.37}$$

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 + \cdots \tag{2.38}$$

with

$$\mathbf{D}_j = \frac{q_j}{4\pi r_j^2}\hat{\mathbf{r}}_j \tag{2.39}$$

$$\mathbf{E} = \frac{q_j}{4\pi\epsilon_0 r_j^2}\hat{\mathbf{r}}_j \tag{2.40}$$

For a continuous distribution of charge where the charge is broken up into a very large number of infinitesimal elements, the electric field at P is the vector sum of the fields $\Delta\mathbf{E}$, due to all the elements Δq of the charge distribution (see Figure 2.3); thus, from equation (2.3), we derive

$$\Delta\mathbf{E} = \frac{\Delta q}{4\pi\epsilon r^2}\hat{\mathbf{r}} \tag{2.41}$$

The total electric field strength at P is

$$\mathbf{E} = \frac{1}{4\pi\epsilon}\int_{charge\ distribution} \frac{dq}{r^2}\hat{\mathbf{r}} \tag{2.42}$$

© 2015 by Taylor & Francis Group, LLC

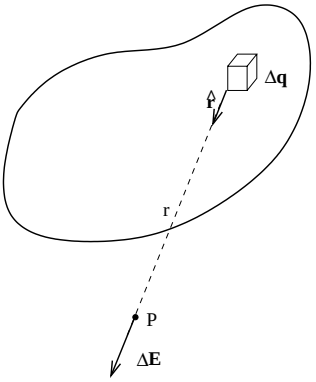


FIGURE 2.3: Illustration of Gauss' law.

2.1.6 Gauss' Law for Electrostatics

The electric flux, Φ_E, through the surface is defined as the product of the area, $\mathtt{S}$, and the normal component of the electric field, $E_\perp (= E\cos\theta)$, in which θ is the angle between $\mathbf{E}$ and a perpendicular to the surface (see Section 2.1.4). On integrating over such a surface, $\mathtt{S}$, the total charge on the inner sphere by the charge Q is

$$\Phi_E = \oint_{\mathtt{S}} \mathbf{D}\cdot d\mathbf{S} = Q \tag{2.43}$$

since $\Phi_E = Q$, measured in coulombs, where Q is the charge enclosed by $\mathtt{S}$, Q can represent either a positive or a negative quantity. The enclosed charge

$$Q = \int_{\mathtt{V}} \rho d\mathtt{V} \tag{2.44}$$

For a spherical volume the integral is

$$\int_{\mathtt{V}} \rho d\mathtt{V} = \int 4\pi r^2 \rho dr \tag{2.45}$$

In the following, we enumerate some cases of Gauss' law:

1. Charged spherical shell: Let us consider a charge Q, distributed uniformly in a thin spherical shell of radius R. Here, we distinguish two cases:

   (a) A point outside the shell: Let us consider a sphere $\mathtt{S}$, of radius r centered at $\mathtt{O}$. The electric field strength at a point $\mathtt{P}$, which is located outside the shell of radius R, i.e., $r > R$, is given by (see Figure 2.4)

   $$\mathbf{E}(\mathbf{r}) = \frac{Q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}} \tag{2.46}$$

© 2015 by Taylor & Francis Group, LLC

which is the same as if all the charges were concentrated at the center of the shell.

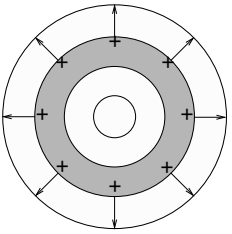


FIGURE 2.4: Charged metal sphere.

(b) A point inside the shell: Let us consider a spherical surface S, of radius r, centered on O, but inside the shell of radius R, i.e., $r < R$. In this case, the electric field strength is

$$\mathbf{E}(\mathbf{r}) = 0, \tag{2.47}$$

which means that there is no electric field at any point inside a closed charged shell.

2. Spherical charge distribution: Let us consider that the electric field strength due to a charged sphere, in which the charge is uniformly distributed throughout its volume, varies with distance from the center of the sphere. Let Q be the total charge that is distributed uniformly throughout the sphere of radius a, i.e., the density of the charge, $\frac{Q}{4\pi a^3/3}$, is everywhere inside the sphere. The electric field strength is given by

$$\mathbf{E}(\mathbf{r}) = \begin{cases} \dfrac{Q}{4\pi\epsilon_0 a^3}\hat{\mathbf{r}} & \text{if } 0 < r < a \\ \dfrac{Q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}} & \text{if } r > a \end{cases} \tag{2.48}$$

3. Field of a line charge: For an infinitely long line of charge (see Figure 2.5a) with a linear charge density λ, the magnitude of the field, dE, is

$$dE = \frac{1}{4\pi\epsilon_0}\frac{dQ}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{\lambda dy}{x^2+y^2}$$

with $r = \sqrt{x^2+y^2}$.

Since the electric field is a vector, we cannot add the magnitude of the field from the whole ring together. By invoking symmetry to cancel the y components of the field, we may write

$$\begin{aligned} E &= E_x = \int \cos\theta dE \\ &= \frac{\lambda}{4\pi\epsilon_0}\int_{-\infty}^{\infty}\frac{\cos\theta dy}{x^2+y^2} \end{aligned} \tag{2.49}$$

© 2015 by Taylor & Francis Group, LLC

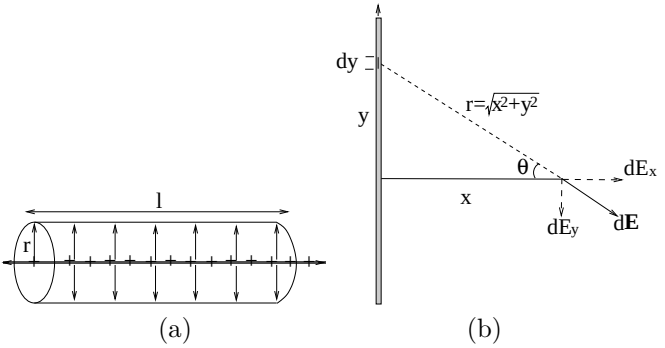


FIGURE 2.5: (a) Using Gauss' law to find the field of a line charge and (b) a line of charge.

where θ is the angle the vector field of dQ makes with the x direction.

In order to find the field at a fixed distance, x, we need to convert the integral in terms of either θ or y, since θ and y are not constant; here, θ and y are not independent, $y = \tan\theta$. Hence, we need to find an expression for dy (see Figure 2.5b) in terms of x and θ:

$$\frac{dy}{d\theta} = \frac{x}{\cos^2\theta}$$

The integral becomes

$$\begin{aligned} E &= \frac{\lambda}{4\pi\epsilon_0}\int_{-\infty}^{\infty}\frac{\cos\theta dy}{x^2+y^2} = \frac{\lambda}{4\pi\epsilon_0}\int_{-\pi/2}^{\pi/2}\frac{x\cos\theta d\theta}{\cos^2\theta(x^2+x^2\tan^2\theta)} \\ &= \frac{\lambda}{4\pi\epsilon_0}\int_{-\pi/2}^{\pi/2}\cos\theta d\theta = \frac{\lambda}{2\pi\epsilon_0 x} \end{aligned} \tag{2.50}$$

4. Field due to a ring of charge: Let us consider that in a ring (see Figure 2.6) of charge with radius a, where the total charge is Q. The field is at a point on the axis going through the center of the ring a distance x, from the center of the ring. Therefore, the magnitude of the field due to this element is

$$dE = \frac{1}{4\pi\epsilon_0}\frac{dQ}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{dQ}{x^2+a^2}$$

in which $dQ = \frac{Q}{2\pi a}dl$ is a charge element and dl a length element.

Invoking symmetry makes the problem easy because all the $dE_\perp$ com-

© 2015 by Taylor & Francis Group, LLC

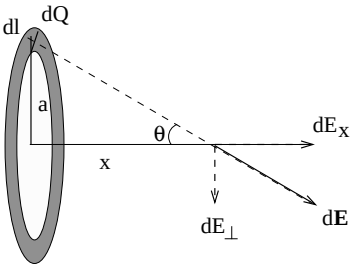


FIGURE 2.6: Field in a ring of charge.

ponents cancel; therefore, the electric field is derived as

$$\begin{aligned} E &= \int dE_x = \int \cos\theta dE \\ &= \int_0^{2\pi a} \frac{1}{4\pi\epsilon_0}\frac{Q}{2\pi a}\frac{\cos\theta dl}{x^2+a^2} \\ &= \frac{Q}{4\pi\epsilon_0}\frac{x}{\sqrt{x^2+a^2}} \end{aligned} \tag{2.51}$$

5. Field due to a uniform disk of charge: In order to find the field due to a uniform disk of charge (see Figure 2.7), which has an areal charge density σ, we can use the result for a ring of charge. Let us consider the field as an integral over a set of these rings. On taking a ring of charge dQ

$$dE = \frac{1}{4\pi\epsilon_0}\frac{zdQ}{(x^2+r^2)^{3/2}} \tag{2.52}$$

On relating the charge element $dQ(=\sigma 2\pi r dr)$ to the width of the ring, dr, we find

$$dE = \frac{z\sigma r dr}{2\epsilon_0(x^2+r^2)^{3/2}} \tag{2.53}$$

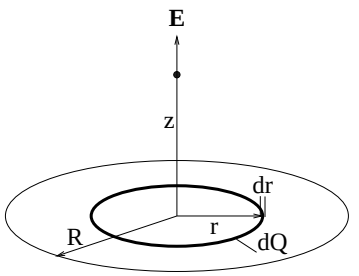


FIGURE 2.7: Field due to a uniform disk of charge.

© 2015 by Taylor & Francis Group, LLC

and therefore, the field is derived as

$$\begin{aligned} E &= \frac{z\sigma}{2\epsilon_0}\int_0^R \frac{r}{(x^2+r^2)^{3/2}}dr \\ &= \frac{\sigma}{2\epsilon_0}\left[1-\frac{z}{(x^2+r^2)^{3/2}}\right] \end{aligned} \tag{2.54}$$

It is pertinent to note that $E = \sigma/(2\epsilon_0)$ is valid for any infinite plane of charge for the limit as $R \to \infty$ or when $z \ll R$.

6. Field of an infinite flat sheet of charge: Smooth distribution of electric charge in a thin sheet is called a surface charge distribution. Consider a flat sheet infinite in extent, with the constant surface charge density σ. The electric field on either side of the sheet, irrespective of its magnitude, should point perpendicular to the plane of the sheet. The field has the same magnitude and the opposite direction at two points, P and P′, equidistant from the sheet on opposite sides. Consider a cylindrical Gaussian surface (see Figure 2.8) of cross-section area S, which has the flat surfaces that have flux passing through them, that is,

$$\begin{aligned} \oint \mathbf{E}\cdot d\mathbf{S} &= -E(-\mathtt{S}) + E\mathtt{S} = E(2\mathtt{S}) \\ &= \frac{Q}{\epsilon_0} = \frac{\sigma \mathtt{S}}{\epsilon_0} \end{aligned} \tag{2.55}$$

The charge enclosed is $\sigma\mathtt{S}/\epsilon_0$ or

$$E = \frac{\sigma}{2\epsilon_0} \tag{2.56}$$

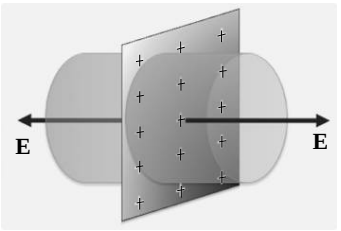


FIGURE 2.8: Infinite plane of charge.

The field strength is independent of the distance from the sheet r. The field of an infinitely long line charge varies inversely as the distance from the line, while the field of an infinite sheet has the same strength at all distances. These are consequences of the fact that the field of a point

© 2015 by Taylor & Francis Group, LLC

charge varies as the inverse square of the distance. In the following we elucidate the field inside and outside of two parallel sheets of charge with opposite signs.

(a) Field inside a conductor: If we consider a static situation inside a conductor (see Figure 2.9a), the field becomes zero. It is due to the fact that if there is an electric field, the free electrical charges would move in response to the field because they experience a force. If there is a net charge on a metal, it arranges itself so that the field inside is equal to zero. The result is that the charges are arranged on the surface of the conductor. The electric fields lines are always perpendicular to a conducting surface; any component of the field parallel to the surface would lead to a movement of charge along the surface.

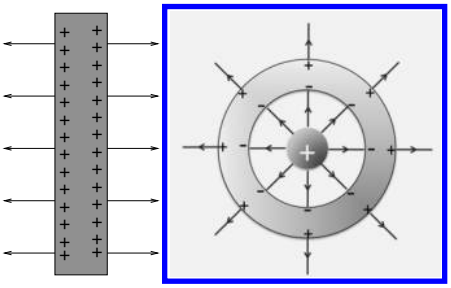


FIGURE 2.9: Electric field (a) inside a conductor and (b) inside a container.

(b) Charge inside a conducting container: If we consider a charge inside a conducting sphere, the inside surface of the sphere should develop a charge to stop the field from developing inside the conductor. In order to maintain neutrality, the outer sphere should develop an equal and opposite charge as displayed in Figure 2.9b. The field outside the sphere is the same as if the sphere is not there at all.

2.1.7 Potential Difference

The electric potential defines the electric field in electrostatics. If a test charge is moved around any closed path in a static electric field, no work is done. Since the charge returns to its starting point, the forces encountered on one part of the path are exactly offset by opposite forces on the remainder of the path. The mathematical statement for conservation of energy in a static electric field is

$$\oint \mathbf{E}\cdot d\mathbf{l} = 0 \tag{2.57}$$

© 2015 by Taylor & Francis Group, LLC

where $d\mathbf{l}$ is the vector differential length.

However, this statement is not true for time-varying fields, in which case, Faraday's law (see Section 2.3.1) applies. The potential difference (voltage) between two points, $\mathtt{P}_1$ and $\mathtt{P}_2$, in an electric field, E, is defined as the work required to move a unit positive test charge from $\mathtt{P}_1$ to $\mathtt{P}_2$, and is given by

$$V_{21} = -\int_{\mathtt{P}_1}^{\mathtt{P}_2} \mathbf{E}\cdot d\mathbf{l} = \mathcal{E} \qquad \text{J.C}^{-1} \text{ or V} \tag{2.58}$$

with $\mathcal{E}$ as the electromotive force (EMF), which is defined to be the limiting maximum potential difference between the terminals of a voltage generator (e.g., battery) as the current drawn from it is reduced towards zero.

The potential difference is independent of the path taken from $\mathtt{P}_1$ to $\mathtt{P}_2$, i.e., it depends on the endpoints. The negative sign in equation (2.58) indicates that the field does work on the positive charger in moving from $\mathtt{P}_1$ to $\mathtt{P}_2$ and there is a fall in potential. With the choice of zero potential, we introduce an electric potential function, $V(r)$, in which r is the distance from the point-like charged object with charge Q

$$V(r) = \frac{1}{4\pi\epsilon_0}\frac{Q}{r} \tag{2.59}$$

In the presence of more than one point charge, the electric potential becomes the sum of potentials due to individual charges, that is,

$$V(r) = \frac{1}{4\pi\epsilon_0}\sum_{j=1}^{N}\frac{q_j}{r} \tag{2.60}$$

2.1.7.1 Deriving Electric Field from Electric Potential

In electrostatics, to avoid the addition of vectors when applying Coulomb's law (see Section 2.1.1) to get the electric field intensity at a point, we develop the concept of a potential and obtain the field as the gradient (see Appendix B.2.1) of this potential function since potential produced by an electric charge is a scalar quantity. In equation (2.57), we find the relation between the electric field $\mathbf{E}$ and the electric potential V. Let us consider that two points are separated by a small distance, $d\mathbf{l}$; hence, the following differential form emerges:

$$d\mathtt{V} = -\mathbf{E}\cdot d\mathbf{l} \tag{2.61}$$

Here, $\mathbf{E} = \hat{\mathbf{x}}E_x + \hat{\mathbf{y}}E_y + \hat{\mathbf{z}}E_z$ and $d\mathbf{l} = \hat{\mathbf{x}}dx + \hat{\mathbf{y}}dy + \hat{\mathbf{z}}dz$ in Cartesian coordinates, and therefore,

$$\begin{aligned} d\mathtt{V} &= \left(\hat{\mathbf{i}}E_x + \hat{\mathbf{j}}E_y + \hat{\mathbf{k}}E_z\right)\cdot\left(\hat{\mathbf{i}}dx + \hat{\mathbf{j}}dy + \hat{\mathbf{k}}dz\right) \\ &= E_x dx + E_y dy + E_z dz \end{aligned} \tag{2.62}$$

© 2015 by Taylor & Francis Group, LLC

On defining directional derivatives $\partial V/\partial x$, $\partial V/\partial y$, and $\partial V/\partial z$, we may construct at each point in space a vector, whose x, y, z components are equal to the respective partial derivatives:

$$d\mathtt{V} = \frac{\partial V}{\partial x}dx + \frac{\partial V}{\partial y}dy + \frac{\partial V}{\partial z}dz \tag{2.63}$$

therefore,

$$E_x = -\frac{\partial V}{\partial x};\ E_y = -\frac{\partial V}{\partial y};\ E_z = -\frac{\partial V}{\partial z}$$

Let us introduce a differential quantity, called the Del operator, $\boldsymbol{\nabla}$ (see equation B.9; Appendix B.1.3). The dot product of the gradient of a function with a unit vector gives the slope of the function in the direction of the unit vector. The electric field is recast as

$$\mathbf{E} = -\left(\hat{\mathbf{i}}\frac{\partial V}{\partial x} + \hat{\mathbf{j}}\frac{\partial V}{\partial y} + \hat{\mathbf{k}}\frac{\partial V}{\partial z}\right) = -\boldsymbol{\nabla}V \tag{2.64}$$

The differential operator, $\boldsymbol{\nabla}$, operates on a scalar quantity (electric potential) and results in a vector quantity (electric field). If the charge distribution possesses spherical symmetry, the resulting electric field is a function of the radial distance, r, that is

$$\mathbf{E} = -\left(\frac{d\mathtt{V}}{dr}\right)\hat{\mathbf{r}} \tag{2.65}$$

for instance, the electric potential due to a point charge Q is $V(r) = Q/4\pi\epsilon_0 r$, and ϵ_0 the permittivity in free space. Therefore, the electric field is expressed as

$$\mathbf{E} = \frac{Q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}} \tag{2.66}$$

In a 2-dimensional system, the curves characterized by constant $V(x, y)$ are called the equipotential curves. In 3D, surfaces such that $V(x, y, z) =$ constant are called the equipotential surfaces. Since $\mathbf{E} = -\nabla V$, the direction of $\mathbf{E}$ at a point is always perpendicular to the equipotential through that point. The properties of equipotential surface are

- the electric field lines are perpendicular to the equipotentials and point from higher to lower potentials;
- the equipotential surfaces produced by a point charge form a family of concentric spheres and, for constant electric field, a family of planes perpendicular to the field lines;
- the tangential component of the electric field along the equipotential surface is zero; and
- no work is needed to move a particle along an equipotential surface.

© 2015 by Taylor & Francis Group, LLC

As stated in Section 2.1.3.3, the electric field caused by a continuous charge distribution $\rho(\mathbf{r})$ is given by

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int_{\mathtt{V}} \rho(\mathbf{r}')\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}d\mathtt{V}' \tag{2.67}$$

where $d\mathtt{V}'$ is the volume element and $\rho(\mathbf{r},t)$ the local charge density at the position vector $\mathbf{r}(=x,y,z)$.

A given charge distribution creates the electrostatic potential scalar field, the gradient of which is the electric field. The electrostatic potential $\varphi(\mathbf{r})$, which gives rise to by a continuous charge distribution $\rho(\mathbf{r})$, is

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int_{\mathtt{V}} \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathtt{V}' \tag{2.68}$$

By differentiating equation (2.67), the electric field intensity is derived as the gradient of a potential

$$\mathbf{E}(\mathbf{r}) = -\boldsymbol{\nabla}\int_{\mathtt{V}} \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathtt{V}' = -\boldsymbol{\nabla}\varphi(\mathbf{r}) \tag{2.69}$$

$$= -\left(\hat{\mathbf{i}}\frac{\partial\varphi}{\partial x} + \hat{\mathbf{j}}\frac{\partial\varphi}{\partial y} + \hat{\mathbf{k}}\frac{\partial\varphi}{\partial z}\right) \tag{2.70}$$

Equation (2.69) for the electric field is provided in terms of the potential gradient, which implies that the electric vector field is always normal to the equipotential surface and is directed along the direction of the decreasing potential.

2.1.7.2 Divergence of the Electric Field

The scalar product of the $\boldsymbol{\nabla}$ operator with a vector field is called the divergence of the field. Physically, the divergence is often thought of as the derivative of the net flow of the vector field out of the point at which the divergence is evaluated. The electric field has a definite direction and magnitude at each point. It is written as $E(\mathbf{r})$, with $\mathbf{r}(=x,y,z)$ as the position vector. The divergence (see Appendix B.2.2) of the electric field in Cartesian coordinates is defined as

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \tag{2.71}$$

The divergence of a vector field provides a measurement of whether it acts as a source or sink of flux. An equation with surface enclosing infinitesimal volumes leads to Gauss' divergence theorem:

$$\oint_{\mathtt{S}} \mathbf{E}\cdot d\mathbf{S} = \int_{\mathtt{V}} \boldsymbol{\nabla}\cdot\mathbf{E}\,d\mathtt{V} = \frac{Q}{\epsilon_0} \tag{2.72}$$

© 2015 by Taylor & Francis Group, LLC

therefore, the differential form of Gauss' electric law is derived as

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{\rho}{\epsilon_0}. \tag{2.73}$$

This law (equation 2.73) tells us that wherever space is free of charge, the divergence of the field is zero. The divergence of $\mathbf{E}$ can be derived from equation (2.22)

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{1}{4\pi\epsilon_0}\int \boldsymbol{\nabla}\cdot\left(\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}\right)\rho(\mathbf{r}')d\mathtt{V}' \tag{2.74}$$

Since

$$\boldsymbol{\nabla}\cdot\left(\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}\right) = 4\pi\delta^3(\mathbf{r})$$

the divergence of $\mathbf{E}$ can be recast as

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{\rho(\mathbf{r})}{\epsilon_0} \tag{2.75}$$

which is Gauss' law in differential form (equation 2.73). Here the delta function[2] is $\delta^3(\mathbf{r}) = \delta(x)\delta(y)\delta(z)$ in three dimensions and $\mathbf{r}(= x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}})$ is the position vector, extending from the origin to the point (x, y, z).

The integral form of equation (2.72) can be written as

$$\int_{\mathtt{V}} \boldsymbol{\nabla}\cdot\mathbf{E}d\mathtt{V} = \oint_{\mathtt{S}} \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0}\int_{\mathtt{V}} \rho d\mathtt{V} = \frac{Q}{\epsilon_0} \tag{2.76}$$

2.1.7.3 Poisson and Laplace Equations

The flux of the field intensity produced by a point charge, according to Gauss' law for the electric field, is expressed in the form

$$\oint_{\mathtt{S}} \mathbf{E}\cdot\hat{\mathbf{n}}d\mathtt{S} = 4\pi q \tag{2.77}$$

where $\hat{\mathbf{n}}$ is an outward-drawn unit normal to the element of surface area $d\mathtt{S}$.

The electric charge density ρ is equal to the total charge q in a volume $\mathtt{V}$ divided by the volume; thus, $\rho = q/\mathtt{V}$. Equation (2.77) for a continuous charge distribution can be written as

$$\oint_{\mathtt{S}} \mathbf{E}\cdot\hat{\mathbf{n}}d\mathtt{S} = 4\pi\int_{\mathtt{V}} \rho(\mathbf{r})d\mathtt{V} \tag{2.78}$$

[2]The three-dimensional (3D) delta function is zero everywhere except at $(0, 0, 0)$, where it explodes[130]. Its volume integral is 1:

$$\int_{\text{all space}} \delta^3(\mathbf{r})d\mathtt{V} = \iiint_{-\infty}^{\infty} \delta(x)\delta(y)\delta(z)dxdydz = 1$$

© 2015 by Taylor & Francis Group, LLC

In a medium of free space, by using the differential form of Gauss' law (equation 2.73) and the relation between $\mathbf{E}$ and φ (equation 2.69), the Poisson (S. D. Poisson, 1781−1840) partial differential equation for φ is obtained as[101]

$$\begin{aligned}\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\varphi &= \boldsymbol{\nabla}^2\varphi = \frac{\partial}{\partial x}\left(\frac{\partial\varphi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial\varphi}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{\partial\varphi}{\partial z}\right) \\ &= \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} = -4\pi\rho(\mathbf{r}) \end{aligned} \tag{2.79}$$

where $\boldsymbol{\nabla}\cdot\boldsymbol{\nabla} = \boldsymbol{\nabla}^2$ is the Laplacian operator (see Appendix B.1.4).

Equation (2.79) relates the electric potential $\varphi(\mathbf{r})$ to its electric sources $\rho(\mathbf{r})$ [135]) In regions empty of charge, this equation turns out to be homogeneous, i.e.,

$$\boldsymbol{\nabla}^2\varphi = 0 \tag{2.80}$$

This expression (equation 2.80) is known as the Laplace equation.

2.1.7.4 Curl of the Electric Field

As in Section 2.1.7.2, let us calculate the curl (see Appendix B.2.4) of the electric field produced by a point charge (see equation 2.46). The magnitude of $\boldsymbol{\nabla}\times\mathbf{E}$ is the limiting value of circulation per unit area, that is,

$$(\boldsymbol{\nabla}\times\mathbf{E})\cdot\hat{\mathbf{n}} \equiv \lim_{\Delta\mathtt{S}\to 0}\left[\frac{1}{\Delta\mathtt{S}}\oint_{\mathtt{C}}\mathbf{E}\cdot d\mathbf{l}\right] \tag{2.81}$$

where the right side is a line integral around an infinitesimal region of area, $d\mathtt{S}$, that is allowed to shrink to zero via a limiting process and $\hat{\mathbf{n}}$ is the unit normal vector to this region.

Looking at the radially directed electric field lines produced by the point charge, we can deduce that the curl of an electric field is zero. Taking a line integral of $\mathbf{E}$ from $\mathtt{P}_1$ to $\mathtt{P}_2$ in spherical coordinate system, the line element of integration is

$$d\mathbf{l} = \hat{\mathbf{r}}dx + \hat{\boldsymbol{\theta}}rd\theta + \hat{\boldsymbol{\phi}}r\sin\phi d\phi \tag{2.82}$$

The result is

$$\int_{\mathtt{P}_1}^{\mathtt{P}_2}\mathbf{E}\cdot d\mathbf{l} = \frac{Q}{4\pi\epsilon_0}\int_{\mathtt{P}_1}^{\mathtt{P}_2}\frac{1}{r^2}dr = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \tag{2.83}$$

If the integral is a closed loop integration ($r_1 = r_2$), it turns out to be equation (2.57). According to Stokes theorem (see Appendix B.3) and using the vector definition for the curl, the line integral of a vector function around a closed contour is equal to the integral of the normal component of the curl of that vector function over any surface having the contour ($\mathtt{C}$) as its boundary edge, that is,

$$\int_{\mathtt{S}}(\boldsymbol{\nabla}\times\mathbf{E})\cdot\hat{\mathbf{n}}d\mathtt{S} = \oint_{\mathtt{C}}\mathbf{E}\cdot d\mathbf{l} = 0 \tag{2.84}$$

© 2015 by Taylor & Francis Group, LLC

Here, the components of the vector field $\mathbf{E}$ are continuous and have first partial derivatives within and on the boundary of $\mathtt{C}$; by incorporating the limit of $\oint \mathbf{E}{\cdot}d\mathbf{l}$ as the area, $\Delta\mathtt{S}$ approaches zero. The equations for static electric fields are obtained by neglecting the derivative with respective to time; that is, there is no variation of field with respect to time or a constant current is flowing. Since the curl of a gradient vanishes (see equation B.32; Appendix B.2.4), we get

$$\boldsymbol{\nabla}\times\mathbf{E} = 0 \tag{2.85}$$

which implies that electrostatic forces are conserved forces and

$$\boldsymbol{\nabla}{\cdot}\mathbf{D} = \rho \tag{2.86}$$

From the electrostatic equations (2.85 and 2.86), we envisage that the circulation of the electric field is zero and the scalar source for the electric charge density is ρ.

2.1.8 Field from Line and Surface Charges

The radial electric field from a uniform line charge of infinite extent is

$$E = \frac{\lambda}{2\pi\epsilon_0 r} \tag{2.87}$$

in which the unit of the radial electric field strength is $\mathrm{V.m^{-1}}$, λ the line charge density denoted by $\mathrm{C.m^{-2}}$, and r the radial distance, m.

The traverse component of the electric field is zero. To note, the radial component falls off as $1/r$. The electric field normal to a surface of infinite extent having a uniform surface charge density σ is

$$E_n = \frac{\sigma}{\epsilon_0} \tag{2.88}$$

and

$$D_n = \sigma \tag{2.89}$$

where E_n represents the normal electric field, $\mathrm{V.m^{-1}}$; D_n the normal flux density, $\mathrm{C.m^{-2}}$; and σ the surface charge density, $\mathrm{C.m^{-2}}$.

An isolated conductor carrying a charge Q has a certain potential ϕ_0. With zero potential at infinity, Q is proportional to ϕ_0. The constant of proportionality, known as capacitance, C, depends on the size and shape of the conductor, i.e., $Q = C\phi_0$. For an isolated spherical conductor of radius a, ϕ_0 is equal to Q/a. Let us consider that two flat conducting plates with area $\mathtt{A}$ are arranged parallel to each other, separated by a distance s (see Figure 2.10).

Let us suppose that there is a charge q on one plate and $-Q$ on the other, and ϕ_1 and ϕ_2 are the respective potentials. On treating uniformly, its magnitude is $(\phi_1 - \phi_2)/s$. The field due to one of these plates of charges can

© 2015 by Taylor & Francis Group, LLC

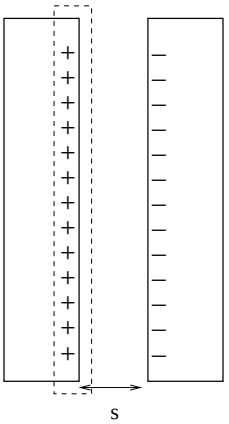


FIGURE 2.10: Two flat conducting plates parallel to one another separated by a distance, s.

be found from Gauss' law $E = Q/\epsilon_0 \mathtt{A}$, and the corresponding density of the surface charge on the inner surface of one of the plates is

$$\sigma = \epsilon_0 E = \frac{\epsilon_0(\phi_1 - \phi_2)}{s} \tag{2.90}$$

On neglecting edge effects, we write an expression for the total charge on one plate as

$$Q = \mathtt{A}\frac{\epsilon_0(\phi_1 - \phi_2)}{s} = \mathtt{A}\frac{\epsilon_0 V}{s} \tag{2.91}$$

with $V(= \phi_1 - \phi_2)$ as the potential difference.

The capacitance, C, the amount of charge separated per volt applied, is $C = Q/V$ and for a parallel plate capacitor is given by

$$C = \frac{\epsilon_0 \mathtt{A}}{s} \tag{2.92}$$

In order to derive the energy stored in a capacitor, let us consider a capacitance C with a potential difference V between the plates, which is equal to the charge, i.e., $Q = V$. There is a charge Q in one plate and $-Q$ on the other. As we increase the charges from Q to $Q + dQ$ by transporting a positive charge dQ from the negative to the positive plate, working against V the work that has to be done is $dW = VdQ = QdQ/C$. The work done in building up the charge separation is given by

$$W = \frac{1}{C}\int_0^Q QdQ = \frac{Q^2}{2C} \tag{2.93}$$

This is equivalent to the amount of energy w stored in the capacitor, which is

© 2015 by Taylor & Francis Group, LLC

expressed as

$$w = \frac{1}{2}CV^2 \tag{2.94}$$

For the parallel plate capacitor with plate area $\mathtt{A}$ and separation s, the capacitance is dictated by equation (2.92) and the electric field is $E = V.s^{-1}$. Hence, equation (2.94) is equivalent to

$$w = \frac{1}{2}\frac{\epsilon_0 \mathtt{A}}{s}V^2 \tag{2.95}$$

2.1.9 Current Density

The flow of electric current, i, in a wire is due to the motion of electrically charged particles (electrons) through the conducting medium. The current is directed opposite to the direction of the electrons. The current through any cross-sectional area is defined as the time rate at which the electric charge passes through the area, that is,

$$i = \frac{dq}{dt} \tag{2.96}$$

where t is the time measured in seconds and $q = \int i dt$.

Assume that the electron density is $-\lambda\,\mathrm{C.m^{-1}}$ and that each electron is moving with a velocity $\mathbf{v}$; hence, in terms of dq and dt, we may write

$$i = \frac{dq}{dt} = \frac{dq}{dl}\frac{dl}{dt} = \lambda v \tag{2.97}$$

where dl is the length of a tiny segment of the wire and λ the linear charge density; the direction of dl is defined to be equal to the direction of the current, i.e., opposite to the direction of the velocity of the electrons.

The current is measured in amperes (A) and any current carrying medium is known as a conductor. Here q is the net charge enclosed by the closed surface $\mathtt{S}$. If a conductor has uniform cross-sectional area, the amount of mobile charge per unit volume of the conductor is the charge density, ρ. The flow of charge through such a conductor at a particular point is given by

$$i = \int_{\mathtt{S}} \mathbf{J}\cdot d\mathbf{S} \qquad \mathrm{A.m^{-2}} \tag{2.98}$$

where $\mathbf{J}(= J_x, J_y, J_z)$ is the electric current density through $\mathtt{S}$ with normal $\mathbf{J} = \mathbf{J}/|\mathbf{J}|$, and $d\mathbf{S}$ is the unit vector of the element perpendicular to the element; the current density describes the magnitude and direction of the amount of current flowing per unit area at a given point inside a material.

The current density at a point relates to the total electric field and the specific conductivity (a measure of how much electrical current flows through an object) of the material, σ, that is,

$$\mathbf{J} = \sigma\mathbf{E} \qquad \mathrm{A.m^{-2}} \tag{2.99}$$

© 2015 by Taylor & Francis Group, LLC

where σ is the charge per unity area when an electric charge is distributed over a surface, and $\mathbf{E}$ is the electric field.

Georg Simon Ohm (1789−1854) observed that the current flowing through most material is proportional to the applied voltage, V, so many (but not all) objects have a resistance defined by $R(= V/i)$ (Ohm law) which is measured in ohms (Ω). The resistance opposes the current causing the dissipation of energy in the form of heat. The resistance of a wire is proportional to its length, l, and inversely proportional to its cross-sectional area, $\mathtt{A}$; it also depends on the material property resistivity, ρ, which is measured in $\Omega\, m$. Here, $R = \rho l/\mathtt{A}$, the reciprocal of the resistivity of a material is called the conductivity, $\sigma = 1/\rho$, which is measured in $(\Omega m)^{-1}$.

2.2 Magnetostatics

Studies of magnetostatics involves a magnetic field, which is analogous to the electric field, $\mathbf{E}$, produced by steady non-time-varying currents. The currents are produced by moving charges undergoing translational motion. A magnetization current is also produced if magnetic dipoles are nonuniformly distributed. In order to realize steady currents, a large number of charges should, in principle, be involved so that collection of charges can be regarded as a continuous fluid. Let us consider two parallel straight wires in which current is flowing. The wires are neutral, and hence, there is no net electric force between them. The flow of an electric current down a conducting wire is due to the motion of electrically charged particles (electrons) through the conducting medium. However, if the current in both wires is flowing in the same direction, the wires are attracted to each other. If the current in one of the wires is reversed, the wires are found to repel each other. The force responsible for the attraction and repulsion is called the magnetic force, $\mathbf{F}_B$. If the current flows in a region with a nonzero magnetic field, each electron experiences a magnetic force. Consider a tiny segment of the wire of length dl. The instantaneous force on the infinitesimal current element is

$$d\mathbf{F}_B = id\mathbf{l}\times\mathbf{B} \tag{2.100}$$

If an external magnetic field $\mathbf{B}$ exerts a force on a current-carrying conductor of length l (see Figure 2.11), the force, $\mathbf{F}_B$, formula is written as

$$\mathbf{F}_B = i\mathbf{l}\times\mathbf{B} \qquad \text{N} \tag{2.101}$$

and the magnitude is

$$F_B = ilB\sin\theta \tag{2.102}$$

in which $\mathbf{l}$ is a length vector and the unit of magnetic field, $\mathbf{B}$, is tesla (T), named after Nikola Tesla.

© 2015 by Taylor & Francis Group, LLC

The magnetic force is proportional to the magnitude of the current, the length of the conductor, the strength of the magnetic field and the sine of the angle between the current and the field.

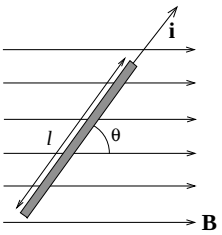


FIGURE 2.11: Magnetic field vs current.

Equation (2.101) states that the force, $\mathbf{F}_B$, is perpendicular to both $\mathbf{l}$ and $\mathbf{B}$, in that case this expression can be written as

$$B = \frac{F}{idl} \tag{2.103}$$

i.e., the magnetic flux density $|\mathbf{B}|$ is defined as a force per unit current element, analogous to the definition of the electric field as a force per unit charge (see equation 2.11). The unit of $|\mathbf{B}|$ is webers per unit area ($\mathrm{Wb.m^{-2}}$), in which $||$ stands for the modulus; the unit of webers is equivalent to a volt-second.

An important difference between the electric field (see equation 2.11) and the magnetic field is that the electric field does work on a charged particle. Since the rate of work, $\mathbf{v}\cdot\mathbf{F} = q\mathbf{v}\cdot\mathbf{E}$, is independent of $\mathbf{B}$, the magnetic force, $\mathbf{F}_B$, exerted on the particle is perpendicular to its velocity $\mathbf{v}$. Therefore, no work is done on the particle which means that its kinetic energy (energy possessed by a body by virtue of its motion), and hence $|\mathbf{v}|$, is constant. The movement of the charge is also equivalent to a current movement, $i\mathbf{l} = q\mathbf{v}$. Thus, the charges experience the magnetic part of the Lorentz force in a constant uniform magnetic field which is expressed as

$$\mathbf{F}_B = q\mathbf{v}\times\mathbf{B} \qquad \text{newtons (N)} \tag{2.104}$$

The force on an infinitesimal current element, $id\mathbf{l}$, immersed in a magnetic field, $\mathbf{B}$, follows from equation (2.104), the definition of current given in equation (2.96), and the velocity, $\mathbf{v} = d\mathbf{l}/dt$. We have

$$\mathbf{v}dq = \frac{d\mathbf{l}}{dt}dq = \frac{dq}{dt}d\mathbf{l} = id\mathbf{l}$$

The force exerted by the magnetic field on a single electron is

$$d\mathbf{F}_e = -e\mathbf{v}\times\mathbf{B} \tag{2.105}$$

© 2015 by Taylor & Francis Group, LLC

2.2.1 Magnetic Flux

The magnetic field **B** around a path of radius R enclosing the wire is given by

$$\begin{aligned} \oint_{\mathtt{C}} \mathbf{B}\cdot d\mathbf{l} &= \int_{\mathtt{S}} \boldsymbol{\nabla}\times\mathbf{B}\cdot d\mathbf{S} \\ &= \frac{\mu_0 i}{2\pi R}\int_{\mathrm{S}} d\mathbf{S} = \mu_0 i \end{aligned} \tag{2.106}$$

where $\int_{\mathtt{S}} d\mathbf{S} = 2\pi R$, μ_0 is the permeability in free space, i the current, and $d\mathbf{l}$ an infinitesimally small length element which points in the direction of the path around the edge.

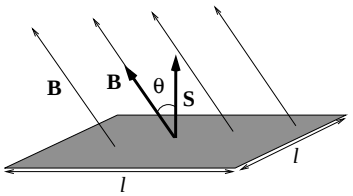


FIGURE 2.12: Magnetic fields passing through an area.

Equation (2.106) may be made independent of the medium by the magnetic field strength, **H**, which is a measure of the intensity of the field whose unit is expressed as A.m^{-1}. Thus, the magnetic flux density, **B**, analogous to the electric flux density, **D**, is defined as

$$\mathbf{B} = \mu_0 \mathbf{H} \tag{2.107}$$

where μ_0 is the permeability in the free space in which the magnetic field acts given in henrys (named after Joseph Henry) per meter (H.m^{-1}).

The lines of magnetic flux are conceptually similar to the lines of electric flux except that the former lines close on themselves, while the latter lines terminate on charges. The magnetic flux Φ_B passing through a surface **S** is defined as the product of the normal magnetic flux density B_n and the surface area $\mathtt{S}$:

$$\Phi_B = \int_{\mathtt{S}} \mathbf{B}\cdot d\mathbf{S} \tag{2.108}$$

When the magnetic field is uniform, the magnetic flux in the example of Figure 2.12 is

$$\Phi_B = \mathbf{B}\cdot\mathbf{S} = BS\cos\theta \tag{2.109}$$

with $A = l^2$.

© 2015 by Taylor & Francis Group, LLC

2.2.2 Ampère's Law

Ampère's law, named after André Marie Ampère (1775−1836), states that the magnetic field that runs around the edge of a closed surface to the current that passes through that surface and is given by equation (2.106) relating to the line integral of **B** around a finite closed path. This is the magnetic field analog of Gauss' law, except that it involves a closed contour rather than a closed surface. Comparing equation (2.106) with equation (2.85) reveals that the static electric field is a conservative field, while the magnetic field is not.

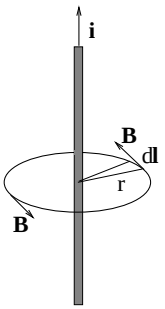


FIGURE 2.13: Field due to a current carrying wire.

For a current carrying wire (see Figure 2.13), the right-hand rule suggests to choose a circular loop centered on the wire, in which case

$$\oint \mathbf{B}\cdot d\mathbf{l} = \oint B dl \tag{2.110}$$

Since **B** is parallel to $d\mathbf{l}$, and as the path is circular,

$$\oint B dl = B(2\pi r) = \mu_0 i \tag{2.111}$$

which leads to

$$B = \frac{\mu_0 i}{2\pi r} \tag{2.112}$$

2.2.3 Biot−Savart Law

Let us assume a current filament as a source. In 1820, Jean Baptiste Biot and Felix Savart announced, based on their experiments, a formula for determining the infinitesimally small magnetic field $d\mathbf{B}$, produced at a point P by an infinitesimal segment of current idl, and this is known as the Biot−Savart Law. The source of static magnetic fields is a charge moving at a constant velocity, namely, direct current (DC; also referred to as steady current or stationary current). The current gives rise to a magnetic field. The differential

© 2015 by Taylor & Francis Group, LLC

magnetic field, $d\mathbf{B}$, is given by

$$d\mathbf{B}(\mathbf{r}) = \frac{\mu_0 i}{4\pi} \frac{d\mathbf{l}\times\hat{\mathbf{r}}}{|\ \mathbf{r}\ |^2} \qquad \text{A.m}^{-1} \tag{2.113}$$

in which $\mathbf{r}$ is the displacement vector from the element $d\mathbf{l}$ to the point and $\hat{\mathbf{r}}$ is a unit vector along $\mathbf{r}$.

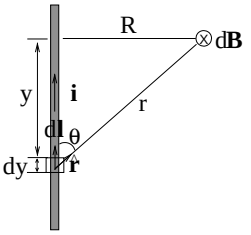


FIGURE 2.14: Illustration of the Biot−Savart law.

Integrating over the entire length of the current element, we get the total magnetic field

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0 i}{4\pi} \oint \frac{d\mathbf{l}\times\hat{\mathbf{r}}}{|\ \mathbf{r}\ |^2} \qquad \text{A.m}^{-1} \tag{2.114}$$

where the circle on the integral sign indicates that the integration is to be performed over the entire length of the closed circuit or loop in which the current, i, flows and $\mathbf{i}$ the vector current.

Equation (2.113) is known to be one form of the Biot−Savart law, which is also attributed to Ampère. The magnetic field due to an infinite straight wire (see Figure 2.14) can be deduced as

$$\begin{aligned} \mathbf{B} &= \int d\mathbf{B} = \frac{\mu_0 i}{4\pi} \int \frac{d\mathbf{l}\times\hat{\mathbf{r}}}{|\ \mathbf{r}\ |^2} \\ &= \frac{\mu_0 i}{4\pi} \int_{-\infty}^{\infty} \frac{\sin\theta dy}{|\ \mathbf{r}\ |^2} \end{aligned} \tag{2.115}$$

with $r^2 = R^2 + y^2$, $\sin\theta = R/r$; we need to write r in terms of y.

The magnetic field, $\mathbf{B}$, due to a infinite straight wire turns out to be

$$\mathbf{B} = \frac{\mu_0 i}{2\pi R} \tag{2.116}$$

The direction of the magnetic field vector in relation to the direction of current flow is given by the right-hand rule. If the thumb points in the direction of the current flow, the fingers indicate the positive direction of the magnetic field.

© 2015 by Taylor & Francis Group, LLC

2.2.4 Magnetic Potential

The magnetic potential, analogous to the electric potential, provides a mathematical way to define a magnetic field, **B**, in classical electromagnetism. As with the electric potential, it is not directly observable, but the field it describes may be measured. As stated in Section 2.1.7.1, the electric field, **E**, can be expressed as the gradient of a scalar potential function (see equation 2.64). While looking for a similar potential function in magnetostatics, we first note a basic difference between the two types of fields. The magnetic field is produced by a current element (idl), which has magnitude and direction, and is, therefore, a vector quantity. This information about the direction of (idl) must be incorporated in the potential function; hence, we postulate a vector magnetic potential, **A**, which is such that an operation of differentiation on it would enable obtaining the magnetic field intensity, **H**. Unlike the divergence which is scalar, the curl is a vector, which leads to define the vector magnetic potential as

$$\mathbf{H} = \boldsymbol{\nabla} \times \mathbf{A} \tag{2.117}$$

To note, in some books the relation is written as

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A} \tag{2.118}$$

where $\mathbf{B} = \mu_0 \mathbf{H}$, μ_0 is the permeability in free space.

So long as μ_0 is constant and does not depend on the magnitude of **H**, we can use either of the forms to define vector magnetic potential. The enumeration of the magnetic fields produced by steady currents has shown that there are three fundamental quantities of magnetostatics: (i) the current density **J**, (ii) the magnetic field **B**, and (iii) the vector potential **A**, which are related, and if one of them is known, the other two can be derived.

2.2.4.1 Divergence and Curl of B

Using the Biot−Savart law (see Section 2.2.3) for a volume current **J**, the divergence and curl of **B** are respectively given by

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0 \tag{2.119}$$

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J} \tag{2.120}$$

Equations (2.119 and 2.120) are similar to equations (2.86 and 2.85) for electrostatics, respectively. The last equation (2.120) is called Ampère's law in differential form. By replacing the current with the surface integral of **J** over an area bounded by the path of integration of magnetic field **B**, we obtain

$$\oint_{\mathrm{C}} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_{\mathrm{S}} \mathbf{J} \cdot d\mathbf{S} \tag{2.121}$$

By invoking Stokes theorem (see Appendix B.3), we may write equation (2.120) as

$$\int_{\mathrm{S}} (\boldsymbol{\nabla} \times \mathbf{B}) \cdot \hat{\mathbf{n}} d\mathbb{S} = \mu_0 \int_{\mathrm{S}} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathbb{S} \tag{2.122}$$

© 2015 by Taylor & Francis Group, LLC

in which $\hat{\mathbf{n}}$ is the unit normal vector.

Equation (2.122) is known as Ampère's law in integral form.

2.2.4.2 Magnetic Vector Potential

There is a fundamental difference between the electric field and the magnetic field although there is no interaction between static electric and magnetic fields. In general, the circulation of the magnetic field intensity $\mathbf{H}$ is nonzero and the vector source of circulation is the current density $\mathbf{J}$. If $\mathbf{J}$ is known as a function of position, we can solve to obtain $\mathbf{H}$ and $\mathbf{B}$. Following equation (B.29; Appendix B.2.4), the curl of the magnetic field (equation 2.118) is written as

$$\boldsymbol{\nabla}\times\mathbf{B} = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A}) - \boldsymbol{\nabla}^2\mathbf{A} \tag{2.123}$$

The divergence of the vector potential vanishes for static fields, since

$$\boldsymbol{\nabla}\cdot\mathbf{A} = \frac{\mu_0}{4\pi}\int \boldsymbol{\nabla}\left(\frac{1}{\mid \mathbf{r}-\mathbf{r}' \mid}\right)\cdot\mathbf{J}(\mathbf{r}')d\mathbb{V}' = 0 \tag{2.124}$$

To note,

$$\int \boldsymbol{\nabla}'\cdot\left(\frac{\mathbf{J}(\mathbf{r}')}{\mid \mathbf{r}-\mathbf{r}' \mid}\right)d\mathbb{V}' = \oint \frac{\mathbf{J}(\mathbf{r}')}{\mid \mathbf{r}-\mathbf{r}' \mid}d\mathbf{S}' = 0$$

and for steady state current flow,

$$\boldsymbol{\nabla}\cdot\mathbf{J} = 0 \tag{2.125}$$

which is equivalent to the statement that at every point in space, for static magnetic fields, the curl of equation (2.118) turns out to be

$$\boldsymbol{\nabla}\times\mathbf{B} = -\boldsymbol{\nabla}^2\mathbf{A} \tag{2.126}$$

However,

$$\begin{aligned}\boldsymbol{\nabla}^2\mathbf{A} &= \frac{\mu_0}{4\pi}\int \boldsymbol{\nabla}^2\left(\frac{1}{\mid \mathbf{r}-\mathbf{r}' \mid}\right)\mathbf{J}(\mathbf{r}')d\mathbb{V}' \\ &= -\mu_0\mathbf{J}(\mathbf{r})\end{aligned} \tag{2.127}$$

The solution of this Poisson equation (2.127) is expressed as

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\mathbf{r}')}{\mid \mathbf{r}-\mathbf{r}' \mid}d\mathbb{V}' \tag{2.128}$$

which is similar to equation (2.67). This vector quantity is called the magnetic vector potential. For a given, non-time-varying current density distribution $\mathbf{J}(\mathbf{r})$, the magnetic field $\mathbf{B}(\mathbf{r})$ can be calculated from the Biot−Savart law (see equation 2.114) as

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi}\oint \frac{\mathbf{J}(\mathbf{r}')\times(\mathbf{r}-\mathbf{r}')}{\mid \mathbf{r}-\mathbf{r}' \mid^3}d\mathbb{V}' \tag{2.129}$$

with $\mathbf{J}(\mathbf{r})d\mathbb{V}' = \mathbf{i}(\mathbf{r})dl$ in the case of reducing the distributed current to an idealized zero thickness wire.

© 2015 by Taylor & Francis Group, LLC

2.2.5 Lorentz Force

The magnetic force, $\mathbf{F}_B$, can, under appropriate conditions, gives rise to an induced current in a conducting circuit moving through a magnetic field (see equation 2.101). This force can be witnessed in an experiment where a beam of electrons is projected into a chamber by an electron gun[156]. The electric field due to the magnetic force is deduced from equation (2.104):

$$\mathbf{E} = \frac{\mathbf{F}_B}{q} = \mathbf{v}\times\mathbf{B} \tag{2.130}$$

When a point charge q moves in both electric and magnetic fields, $\mathbf{E}$ and $\mathbf{B}$, the total force exerted on the point charge is the sum of the magnetic and electric forces, that is,

$$\begin{aligned} \mathbf{F} &= \mathbf{F}_E + \mathbf{F}_B \\ &= \int_{\mathtt{V}} [\rho\mathbf{E} + \rho\mathbf{v}\times\mathbf{B}]\, d\mathtt{V} \\ &= (\mathbf{E} + \mathbf{v}\times\mathbf{B}) \int_{\mathtt{V}} \rho d\mathtt{V} \\ &= q\,(\mathbf{E} + \mathbf{v}\times\mathbf{B}) \qquad \text{N} \end{aligned} \tag{2.131}$$

in which $\int_{\mathtt{V}} \rho d\mathtt{V} = q$ and the expression for the electric force is given in equation (2.11).

Equation (2.131) is called the Lorentz force law, named after Hendrik A. Lorentz (1853−1928), which describes the resultant force experienced by a particle of charge q moving with velocity $\mathbf{v}$, under the influence of both an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$. The total force at a point within the particle is an applied field together with the field due to charge in the particle itself (self-field). In practical situations, the self-force is negligible; therefore, the total force on the particle is approximately the applied force. The expression (equation 2.131), also referred to as Lorentz force density, provides the connection between classical mechanics and electromagnetism. The Lorentz force is utilized in Hall effect devices, in the focusing and deflection of electron beams in cathode-ray tubes, and in galvanometer movements.

If an electric current, i, passes through a specimen (conductor or semiconductor) placed in a magnetic field, $\mathbf{B}$, a potential proportional to the current and to the magnetic field is developed across the material in a direction perpendicular to both the current and to the magnetic field, which is referred to as the Hall effect[40]. The charge carriers experience a Lorentz force (see equation 2.131), which would deflect them toward one side of the slab. The field required to achieve equilibrium is

$$E_H = v_d B \tag{2.132}$$

where v_d is the drift velocity of the carriers in the conductor.

© 2015 by Taylor & Francis Group, LLC

2.3 Time-Varying Fields

Both the electric field $\mathbf{E}(\mathbf{r}, t)$ and the magnetic field $\mathbf{B}(\mathbf{r}, t)$, in which $\mathbf{r}(= x, y, z)$ is the position vector and t the time, are time-dependent which can be specified at every point in space. In what follows, the fundamental laws governing time-varying fields are enumerated in brief.

2.3.1 Faraday's Law

In 1831, Faraday found experimentally, what is referred to as Faraday's law on induction, that the work done by an electric force in moving a unit positive charge around a closed path is proportional to the time rate of decrease of the flux of magnetic force through any open surface. When a changing magnetic field is linked with a coil, an electromotive force (EMF) is induced in it. This change in magnetic field may be caused by changing the magnetic field strength by moving a magnet towards or away from the coil or by moving the coil into or out of the magnetic field as desired. In other words, the magnitude of the EMF induced in the circuit is proportional to the rate of change of flux. The EMF, $\mathcal{E}$, induced on the wire loop $\mathtt{C}$ by a time-varying magnetic flux is expressed as

$$\mathcal{E} = \oint_{\mathtt{C}} \mathbf{E} \cdot d\mathbf{l} = -\frac{d\Phi_B}{dt} \tag{2.133}$$

where $d\mathbf{l}$ is the vector differential length along $\mathtt{C}$, and Φ_B the magnetic flux linked with the closed loop.

The negative sign in equation (2.133) follows from the Lenz law, named after Heinrich F. E. Lenz (1804−1865), which states that the direction of the induced voltage is such that it tends to produce a current flow that opposes the change of flux. This law is a consequence of Newton's third Law. As soon as the current in the electromagnet gets established, the magnetic field $\mathbf{B}$ becomes steady and the induced electric field disappears. When the circuit or loop is fixed, the above equation translates into

$$\oint_{\mathtt{C}} \mathbf{E} \cdot d\mathbf{l} = -\int_{\mathtt{S}} \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \tag{2.134}$$

In a static field, the RHS of equation (2.134) is zero (see equation 2.57). When Faraday's law is applied to the surface $\Delta\mathtt{S}$, we get

$$\frac{1}{\Delta\mathtt{S}} \oint_{\mathtt{C}} \mathbf{E} \cdot d\mathbf{l} = -\frac{1}{\Delta\mathtt{S}} \int_{\Delta\mathtt{S}} \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \tag{2.135}$$

Using Stokes theorem (see Appendix B.3), we incorporate the limit of $\oint \mathbf{E} \cdot d\mathbf{l}$ as the area $\Delta\mathtt{S}$ approaches zero, and using the vector definition for the

© 2015 by Taylor & Francis Group, LLC

curl on equation (2.81), equation (2.135) turns out to be the curl of **E** normal to $\Delta\mathtt{S}$ at the point around which $\Delta\mathtt{S}$ shrinks to zero. Thus, we write

$$\begin{aligned}\lim_{\Delta\mathtt{S}\to 0}\left[\frac{1}{\Delta\mathtt{S}}\oint_{\mathtt{C}}\mathbf{E}\cdot d\mathbf{l}\right] &= \lim_{\Delta\mathtt{S}\to 0}\left[-\frac{1}{\Delta\mathtt{S}}\int_{\Delta\mathtt{S}}\frac{\partial\mathbf{B}}{\partial t}\cdot d\mathbf{S}\right]\\ &= -\hat{\mathbf{n}}\cdot\frac{\partial\mathbf{B}}{\partial t}\end{aligned} \tag{2.136}$$

Therefore, the line integral of **E** around the periphery of the area is given by

$$\oint_{\mathtt{C}}\mathbf{E}\cdot d\mathbf{l} = \int_{\mathtt{S}}(\boldsymbol{\nabla}\times\mathbf{E})\cdot d\mathbf{S} \tag{2.137}$$

From these two equations (2.136 and 2.137), we obtain

$$\int_{\mathtt{S}}(\boldsymbol{\nabla}\times\mathbf{E})\cdot d\mathbf{S} = -\int_{\mathtt{S}}\frac{\partial\mathbf{B}}{\partial t}\cdot d\mathbf{S} \tag{2.138}$$

Since $d\mathbf{S}$ in equation (2.138) applies to any surface element, it is arbitrary in definition of curl (see Section 2.1.7.4) and therefore, the integrands are equal.

2.3.2 Inductance

Inductance is the resistance of a circuit element to the change of electric current flowing through it. Inductors do this by generating a self-induced electromotive force (EMF) within itself as a consequence of their changing magnetic fields. Arising from Faraday's law (see Section 2.3.1), the inductance L (the symbol is chosen to honor H. Lenz) may be defined in terms of the EMF generated to oppose a given change in current:

$$\mathcal{E} = -L\frac{di}{dt} \tag{2.139}$$

The unit for inductance is the henry (H). Increasing current in a coil of wire generates a counter EMF which opposes the current. Applying the voltage law allows us to envisage the effect of this EMF on the circuit equation. The fact that the EMF always opposes the change in current is an example of the Lenz law.

2.3.2.1 Self-Inductance

Self-inductance L is the ratio of the back EMF ($\mathcal{E}$) to the time rate of change of the current producing it. Mathematically, the self-induced EMF is written as

$$\mathcal{E}_L = -N\frac{d\Phi_B}{dt} = -N\frac{d}{dt}\int\mathbf{B}\cdot d\mathbf{S} \tag{2.140}$$

where N is the number of turns.

© 2015 by Taylor & Francis Group, LLC

The inductance is a function of geometry. For instance, in a solenoid, the magnetic field is concentrated into a nearly uniform field in the center; the field outside is weak and divergent. The magnetic field of a solenoid is

$$\mathbf{B} = \frac{\mu_0 N i}{l}\hat{\mathbf{n}} = \mu_0 n i \hat{\mathbf{n}} \tag{2.141}$$

in which l is the length of the coil, μ_0 the permeability of free space, $n = N/l$ the number of turns per unit length, and hence, for a long coil the EMF is approximated by

$$\mathcal{E} = -\frac{\mu_0 N^2 \mathtt{A}}{l}\frac{di}{dt} \tag{2.142}$$

$\mathtt{A}(\pi R^2)$ is the cross-sectional area and R the radius.

From equation (2.139), we get the self-inductance as

$$L = \frac{N\Phi_B}{l} = \mu_0 n^2 \pi R^2 l \tag{2.143}$$

2.3.2.2 Mutual Inductance

When the EMF is induced into an adjacent component situated within the same magnetic field, the EMF is said to be induced by mutual induction, M; mutual induction is the basic operating principle of transformers, motors, relays, etc. Let us assume that two coils, $\mathtt{C}_1$ and $\mathtt{C}_2$, are placed near each other. The first coil has N_1 turns and carries the current i_1 which gives rise to a magnetic field, $\mathbf{B}_1$. Due to the close proximity, some of the magnetic field lines through $\mathtt{C}_1$ would pass through $\mathtt{C}_2$. Let Φ_{21} demote the magnetic flux through one turn of $\mathtt{C}_2$ due to i_1. By varying i_1 with time, an induced EMF associated with the changing magnetic flux in the second coil ensues, that is,

$$\mathcal{E}_{21} = -N_2\frac{d\Phi_{21}}{dt} = -\frac{d}{dt}\int_{\mathtt{C}_2} \mathbf{B}_1 \cdot d\mathbf{S}_2 \tag{2.144}$$

The time rate of change of magnetic flux, Φ_{21}, in $\mathtt{C}_2$ is proportional to the time rate of change of the current in $\mathtt{C}_1$

$$N_2\frac{d\Phi_{21}}{dt} = M_{21}\frac{di_1}{dt} \tag{2.145}$$

where the proportionality constant $M_{21}(= N_2\Phi_{21}/i_1)$ is called the mutual inductance.

Similarly, if there is a current, i_2, in $\mathtt{C}_2$ and it is varying with time, the induced EMF in $\mathtt{C}_1$ turns out to be

$$\mathcal{E}_{12} = -N_1\frac{d\Phi_{12}}{dt} = -\frac{d}{dt}\int_{\mathtt{C}_1} \mathbf{B}_2 \cdot d\mathbf{S}_1 \tag{2.146}$$

© 2015 by Taylor & Francis Group, LLC

and a current is induced in $\mathtt{C}_1$. This changing flux in $\mathtt{C}_1$ is proportional to the changing current in $\mathtt{C}_2$

$$N_1 \frac{d\Phi_{12}}{dt} = M_{12} \frac{di_2}{dt} \tag{2.147}$$

with $M_{12} = N_1 \Phi_{12}/i_2$.

By using the reciprocity theorem which combines Ampère's law (see Section 2.2.2) and Biot−Savart law (see Section 2.2.3), we get $M_{12} = M_{21} = M$.

2.3.3 Resonant Circuits

Thus far the waves we have considered (see Section 1.2) have been propagated by means of the periodic motions of material particles under the action of a restoring force. Let us extend the idea to the motions of electrons in terms of voltage and current. For a circuit containing resistance, R, the relation between current and voltage is the same as for the direct current (DC). The voltage, V, across it is given by $V = iR$, in which i is the instantaneous current, and if the voltage is varying sinusoidally, we have

$$\begin{aligned} V &= V_0 \sin(\omega t + \psi) \\ i &= i_0 \sin(\omega t + \psi) \end{aligned} \tag{2.148}$$

with $\omega (= 2\pi f)$, f as the frequency of the generator, and ψ as a phase angle which is the same for the current as for the voltage, i.e., the current and the voltage are in phase.

As stated in Section 2.3.2, if a coil is connected across the terminals of an AC generator, an EMF is set up in the coil as the current changes (see equation 2.139). For a circuit containing a coil of inductance, L, the current and applied voltage are related by

$$\begin{aligned} i &= i_0 \sin(\omega t + \psi - \pi/2) \\ \frac{V_0}{i_0} &= \omega L \end{aligned} \tag{2.149}$$

in which V is dictated by equation (2.148).

A generator produces an alternating EMF that establishes an alternating current (AC). In an oscillating LC circuit, the energy shuttles periodically between the electric field of the capacitor and the magnetic field of the inductance. The instantaneous values of two forms of energy are $w_e = q^2/2C$ and $w_m = Li^2/2$, where q is the instantaneous charge on the capacitor and i the instantaneous current through the inductor. The total energy

$$w = w_e + w_m = \frac{q^2}{2C} + \frac{Li^2}{2}$$

remains constant. On differentiating this equation, we get

$$\frac{dw}{dt} = \frac{d}{dt}\left(\frac{q^2}{2C} + \frac{Li^2}{2}\right) = Li\frac{di}{dt} + \frac{q}{C}\frac{dq}{dt} = 0$$

© 2015 by Taylor & Francis Group, LLC

However, the principle of conservation of energy leads to

$$L\frac{d^2q}{dt^2} + \frac{q}{C} = 0 \tag{2.150}$$

which is the differential equation of LC oscillations; the solution of this equation is

$$q = Q_0 \cos(\omega t + \psi) \tag{2.151}$$

where Q_0 is the charge amplitude and ψ the phase constant.

The second derivative of q is

$$\frac{d^2q}{dt^2} = -\omega^2 Q_0 \cos(\omega t + \psi)$$

On substituting the values of q and its second derivative into equation (2.150), we can determine the angular frequency as $\omega = 1/\sqrt{LC}$.

However, in a circuit containing both an inductor and a resistor, the current lags on the voltage by an amount between 0 and $\pi/2$. The relation between current and voltage is

$$\begin{aligned} V &= i_0\sqrt{R^2 + \omega^2 L^2}\sin(\omega t + \psi) \\ i &= i_0 \sin \omega t \end{aligned} \tag{2.152}$$

where the quantity $\sqrt{R^2 + \omega^2 L^2}$ is referred to as the impedance of the circuit.

On differentiating twice equation (2.148) with respect to time, t, we get

$$\frac{d^2V}{dt^2} = -\omega^2 V \quad \text{and} \quad \frac{d^2i}{dt^2} = -\omega^2 i$$

These expressions are identical in form with the equation of simple harmonic motion (see Chapter 3). Another element for the alternating current (AC) circuit is the capacitor, C, where electric charge is able to pile up on the plates of the capacitor. In this case, the relation between current and voltage is given by

$$\begin{aligned} i &= i_0 \sin(\omega t + \psi + \pi/2) \\ \frac{V_0}{i_0} &= \frac{1}{\omega C} \end{aligned} \tag{2.153}$$

in which V is dictated by equation (2.148).

The current in this case leads the voltage by 90°, while that through an inductor lags on the voltage by 90°. If $\psi = \pi/2$, the voltage is a maximum at time $t = 0$, so no current flows. The charge on the capacitor is a maximum and gives rise to a voltage on the capacitor which balances the voltage. As V starts to fall, the charge leaves the capacitor at an increasing rate, and the current becomes a maximum when V is zero. Current continues to flow as V becomes negative and becomes zero as V reaches its maximum in the negative sense.

© 2015 by Taylor & Francis Group, LLC

Resonance in an electrical circuit implies a particular frequency determined by the values of the resistance, inductance and capacitance. There are two basic kinds of resonant circuits:

1. Series resonance circuit: Let us consider an electric circuit containing a resistance, R, a capacitance, C, and an inductance, L, in series (see Figure 2.15a) with an EMF, $\mathcal{E}(= V_R + V_C + V_L)$, in which V_R, V_C, and V_L are the voltages across the resistance, capacitance, and inductance, respectively. As stated in Section 2.1.8, the charge across the conductor varies directly as the potential difference across it, so that $q = CV_C$, in which V_C is capacitive voltage. Two fundamental laws for electric circuits and networks, known as the Kirchhoff laws, which are named after G. R. Kirchhoff (1824−1887), are

   (a) the sum of the potential differences around a closed circuit is zero, i.e., $\sum V_n = 0$ and

   (b) the algebraic sum of the currents flowing into a junction is zero, $\sum i_n = 0$.

   Hence, for a simple closed circuit consisting of a resistance and inductance, we get

$$\mathcal{E} = L\frac{di}{dt} + Ri$$

   where $\mathcal{E}$ is the imposed EMF and i the instantaneous current at time t.

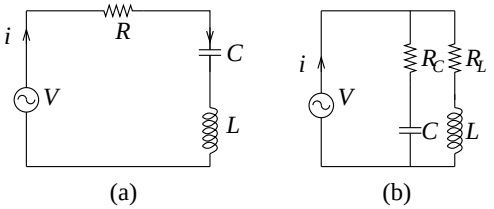


FIGURE 2.15: (a) Series resonance circuit and (b) parallel resonance circuit.

   This expression leads to a linear differential equation of the first order, whose solution is

$$i = \frac{\mathcal{E}}{R}\left(1 - e^{Rt/L}\right) \tag{2.154}$$

   An inspection of equation (2.154) shows that the current increases with increase of t and approaches to the value $\mathcal{E}/R$ as a limit. In general, the R/L is numerically large, so that after a short time, the value of current, i, differs from $\mathcal{E}/R$ by a negligible amount[158]. In the case of a circuit as depicted in Figure 2.15a, by applying the Kirchhoff first law, we get

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int idt = \mathcal{E}_0 \sin \omega t \tag{2.155}$$

© 2015 by Taylor & Francis Group, LLC

in which $i = C\frac{dV_C}{dt}$, $V_C = \frac{1}{C}\int i dt$, and $\mathcal{E}_0$ the amplitude of EMF.

With a resistance, R, present, the total EMF, $\mathcal{E}$, of the circuit (the sum of electric energy and magnetic energy) is no longer constant. It decreases with time as the energy is transferred to thermal energy in the resistance. Therefore, the oscillations of charge, current, and potential difference decrease in amplitude and the oscillations are said to be damped. Substituting dq/dt for i (see equation 2.96) and d^2q/dt^2 for di/dt, we get

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = 0 \tag{2.156}$$

which is a differential equation for damped oscillation in $R\,L\,C$ circuit.

The solution of equation (2.156) is

$$q = Q_0 e^{-Rt/2L}\cos(\omega' t + \psi) \tag{2.157}$$

with $\omega' = \sqrt{\omega^2 - (R/2L)^2}$ and $\omega = 1/\sqrt{LC}$ as the frequency with an undamped oscillation given by equation (2.151).

Equation (2.157) describes a sinusoidal oscillation with an exponential decaying amplitude $Q_0 e^{-Rt/2L}$. The resonance of this series circuit occurs when the inductive and capacitive reactances are equal in magnitude but cancel each other since they are 180° apart in phase; the condition of resonance is characterized by minimum impedance and zero phase. In this circuit, a constant-voltage generator drives a current, i, whose impedance varies with frequency, so that i becomes a function of frequency. The reactance is defined by

$$X_s = X_L - X_C = \frac{V_L}{i} - \frac{V_C}{i} = \omega L - \frac{1}{\omega C}$$

in which V_L is the inductance voltage.

At a particular frequency, called the resonant frequency, $\omega_0 = 1/\sqrt{LC}$, X_s is zero and, hence, i becomes infinite; this resonance phenomenon is known as the series resonance. The resonant frequency is the same as the natural resonant frequency of free oscillation of LC circuit. The impedance, Z, is minimum at resonance, both the current amplitude

$$i = \frac{\mathcal{E}}{Z} = \frac{\mathcal{E}}{\sqrt{R^2 + [\omega L - (1/\omega C)]^2}}$$

and the average power, $\langle P\rangle (= i^2 R)$, would be minimum at resonance.

2. Parallel resonance circuit: In a parallel type circuit (see Figure 2.15b), a constant-current generator supplies a voltage across the circuit whose

© 2015 by Taylor & Francis Group, LLC

impedance varies with frequency. However, here it is the voltage across the circuit which varies. From equations (2.149 and 2.153), we have

$$V_L \sin\left[\omega t - \frac{\pi}{2}\right] = \omega L i \sin \omega t$$
$$V_C \sin\left[\omega t + \frac{\pi}{2}\right] = \frac{i}{\omega C} \sin \omega t \tag{2.158}$$

where ψ is equal to $-\pi/2$ and $+\pi/2$ in the respective equations (2.149 and 2.153); hence,

$$\frac{V_L}{V_C} = -\omega^2 LC \tag{2.159}$$

Here the minus sign indicates that the phases differ by 180°.

2.4 Maxwell's Equations

Maxwell [147] combined together the phenomena in electrostatics, magnetism, and current electricity, discovered by Gauss, Faraday, and Ampère which were considered to be separate and independent, in the form of four differential equations for the electromagnetic field highlighting the relationship between electric and magnetic fields, charge and current. The space and time derivatives of five vectors, namely, the electric vector **E**, the magnetic induction **B**, the electric current density **J**, the displacement current density **D**, and the magnetic vector **H**, are related by Maxwell's equations. To set the stage for the presentation on electromagnetic radiation, the mathematical framework of Maxwell's equations describing the propagation of a radiation field with general time dependence are derived.

2.4.1 Maxwell's First Law

Maxwell's first equation is based on Faraday's law on induction (equation 2.134; see Section 2.3.1), the essence of which is that a time-changing magnetic field induces an electric field. This holds for any medium such as free space, metals, and dielectrics. This is the basis for electromagnetic propagation in free space where electric and magnetic fields produce each other in the absence of any sources or material medium. This is also known as the integral form of Maxwell's equation. Rearranging equation (2.138), we get

$$\int_{\mathrm{S}} \left[(\boldsymbol{\nabla} \times \mathbf{E}) + \frac{\partial \mathbf{B}}{\partial t}\right] \cdot d\mathbf{S} = 0$$

The vanishing of the integral for any arbitrary closed surface requires that the

© 2015 by Taylor & Francis Group, LLC

integrand itself is zero

$$\boldsymbol{\nabla}\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}=0 \tag{2.160}$$

Equation (2.160) is the differential form of Maxwell's equation, which states that the magnitude of the rate of change of the magnetic flux due to a magnetic field through a surface is equal to the magnitude of the electromagnetic force (EMF) along the boundary of the surface. The direction of the EMF opposes the change causing it.

2.4.2 Maxwell's Second Law

Similarly, the differential form for the Ampère−Maxwell law can also be derived. By utilizing the result of the experiments conducted by Ampère, Maxwell deduced a corresponding relation in which the roles of **E** and **H** are interchanged. Introducing equation (2.106) into equation (2.105) yields Ampère's law

$$\oint_{\mathtt{C}}\mathbf{H}\cdot d\mathbf{l}=i \tag{2.161}$$

which relates the line integral of **H** around a finite closed path.

By replacing the current i with the surface integral of the conduction current density $\mathbf{J}_c$ over an area bounded by the path of integration of **H**, we obtain

$$\oint_{C}\mathbf{H}\cdot d\mathbf{l}=\int_{\mathtt{S}}\mathbf{J}_c\cdot d\mathbf{S} \tag{2.162}$$

Let us divide equation (2.162) by the area $\Delta\mathtt{S}$ enclosed by the path and take the limit as $\Delta\mathtt{S}$ approaches zero; therefore, we may write

$$\lim_{\Delta\mathtt{S}\to 0}\left[\frac{1}{\Delta\mathtt{S}}\oint_{\mathtt{C}}\mathbf{H}\cdot d\mathbf{l}\right]=\lim_{\Delta\mathtt{S}\to 0}\frac{\Delta i}{\Delta\mathtt{S}} \tag{2.163}$$

The RHS of equation (2.163) equals the current density at the point around which the area $\Delta\mathtt{S}$ shrinks to zero.

2.4.2.1 Conduction Current Density

Consider a region with a charge density ρ and a current density **J**, which is characterized by its physical constants of conductivity, σ, permeability, μ, and permittivity, ϵ. The presence of an electric field, **E**, in a region with conductivity σ produces a conduction current density

$$\mathbf{J}_c=\sigma\mathbf{E} \tag{2.164}$$

As stated earlier in Section 2.2.2, Ampère's law applies to any closed path and it applies to magnetic fields arising from both conduction and convection currents. In a nonconducting region, a charge density ρ moving with velocity **v** is the equivalent of current density, for instance, electron beams in cathode

© 2015 by Taylor & Francis Group, LLC

ray tubes and electron streams in ionized gases and plasmas. This equivalent current density is known as convection current density $\mathbf{J}_v$ and is defined by

$$\mathbf{J}_v = \rho \mathbf{v} \qquad \text{A.m}^{-2} \tag{2.165}$$

The dimension of ρ is expressed as C.m^{-3}. The total current density, $\mathbf{J}$, is the sum of conduction and convection current densities

$$\mathbf{J} = \mathbf{J}_c + \mathbf{J}_v \tag{2.166}$$

In free space, the volume density of current in a conductor is given by

$$\mathbf{J}(\mathbf{r}, t) = \mathbf{J}_c(\mathbf{r}, t) = \sigma \mathbf{E}(\mathbf{r}, t) \tag{2.167}$$

where $\mathbf{r}(= x, y, z)$ is the 3D position vector.

In equation (2.167), σ is independent of position vector $\mathbf{r}$ and time t, and $\sigma \geq 0$. For a time-dependent current distribution, the surface integral of $\mathbf{J}$ over any closed surface is zero (see equation 2.125). Let us suppose that the current is unsteady, $\mathbf{J}$ being a function of time, t, as well as of space, (x, y, z). Since $\int_{\mathtt{S}} \mathbf{J} \cdot d\mathbf{S}$ is the instantaneous rate at which charge is leaving the enclosed volume, while $\int_{\mathtt{V}} \rho d\mathtt{V}$ is the total charge inside the volume at any instant, we have

$$\int_{\mathtt{S}} \mathbf{J} \cdot d\mathbf{S} = -\frac{d}{dt} \int_{\mathtt{V}} \rho d\mathtt{V} \tag{2.168}$$

The volume charge density ρ and the current density $\mathbf{J}(\mathbf{r}, t)$ are the sources of the electromagnetic radiation. By invoking Stokes theorem (see Appendix B.3), we may write equation (2.162) as

$$\int_{\mathtt{S}} (\boldsymbol{\nabla} \times \mathbf{H}) \cdot \hat{\mathbf{n}} d\mathtt{S} = \int_{\mathtt{S}} \mathbf{J} \cdot \hat{\mathbf{n}} d\mathtt{S} \tag{2.169}$$

The differential form of Ampère's circuital law (see Section 2.2.2) is expressed as

$$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J} \tag{2.170}$$

2.4.2.2 Displacement Current

During the process of charging a capacitor (see Section 2.1.8), Maxwell observed two phenomena:

- a temporary changing current i flows in the wire which produces a magnetic field around the wires, and
- there is no electric current in between the two plates of the capacitors; instead, there is a time-varying electric flux Φ_E.

© 2015 by Taylor & Francis Group, LLC

The question arises whether a magnetic field is induced in the space between the two plates of a capacitor during its charging. On the basis of certain arguments, Maxwell proved that a time-varying electric field **E** must induce a magnetic field **B**, just as time-varying **B** induces **E**. Therefore, Maxwell's law of induction of magnetic field is almost symmetric with equation (2.133) and can be expressed as

$$\oint_{\mathsf{C}} \mathbf{B}\cdot d\mathbf{l} = \mu_0\epsilon_0 \frac{d\Phi_E}{dt} \tag{2.171}$$

where **B** is the magnetic field induced along a closed loop by a changing electric flux Φ_E in the region encircled by that loop, and μ_0 and ϵ_0 are the permeability and permittivity in free space, respectively.

Equation (2.171) says that a changing magnetic field induces an electric field. By combining equations (2.106 and 2.171) into a single equation, the modified version of the Ampère−Maxwell law is obtained as

$$\oint_{\mathsf{C}} \mathbf{B}\cdot d\mathbf{l} = \mu_0 i_c + \mu_0\epsilon_0 \frac{d\Phi_E}{dt} \tag{2.172}$$

where i_c indicates the conduction current encircled by the closed loop, and the product

$$\epsilon_0 \frac{d\Phi_E}{dt} = i_d \tag{2.173}$$

is known as the electric displacement current and has dimension.

2.4.2.3 Ampère's Law with Maxwell Correction

While formulating the equations for the electromagnetic field, Maxwell proposed that in a time-varying field, the time rate of change of the electric flux density is equivalent to a current density and produces a magnetic field similar to that produced by a conduction current density (moving charges in conductors) and a convection current density (moving charges in space). The quantity $\partial\mathbf{D}/\partial t$ is called the displacement current density. When the electric field does not change with time, the displacement current is zero and Ampère's circuital law is applicable. In the presence of both conduction and displacement currents, equation (2.170) can be extended as

$$\begin{aligned} \oint_{\mathsf{C}} \mathbf{H}\cdot d\mathbf{l} &= \int_{\mathsf{S}} \left(\sigma\mathbf{E} + \epsilon\frac{\partial\mathbf{E}}{\partial t}\right)\cdot d\mathbf{S} \\ &= \int_{\mathsf{S}} \left(\mathbf{J}_c + \frac{\partial\mathbf{D}}{\partial t}\right)\cdot d\mathbf{S} \end{aligned} \tag{2.174}$$

The LHS of equation (2.174) represents the magnetomotive force (MMF). The first term on the RHS indicates that MMF originates from current carrying conductor. This was observed by Hans Christian Oersted (1777−1851) [41] and postulated by Ampère. The second term on the RHS indicates that time variation of the flux of electric displacement vector is a source of MMF. This

© 2015 by Taylor & Francis Group, LLC

reflects a symmetry of nature because it shows that according to Faraday's law of electromagnetic induction, variation in the flux of magnetic induction gives rise to electromotive force (EMF) and variation of the flux of electric induction vector gives rise to MMF.

As before, by invoking Stokes theorem (see Appendix B.3), we get the differential form of Maxwell's equation for the time-dependent fields which is derived from Ampère's law

$$\boldsymbol{\nabla}\times\mathbf{H} = \mathbf{J}_c + \frac{\partial \mathbf{D}}{\partial t} \tag{2.175}$$

The second term on the RHS of equation (2.175) was introduced into the equation of electromagnetism by Maxwell, and this is what is referred to as the Ampère−Maxwell equation. It states that the average magnetic field along the boundary of a surface depends on the current per unit area inside the surface and on the rate of change of the electric field across the surface.

2.4.3 Maxwell's Third Law

Maxwell's third law is derived from Gauss' electric law which relates the electric field intensity with the charge distribution producing it. The charge enclosed is equal to the normal component D_n of the flux density over the surface volume, $\Delta\mathtt{V}$, according to Gauss' electric law

$$\oint_{\mathtt{S}} \mathbf{D}\cdot d\mathbf{S} = \int_{\mathtt{V}} \rho d\mathtt{V} \tag{2.176}$$

in which ρ is the volume charge density and $\mathtt{V}$ the volume enclosed by the closed surface, $\mathtt{S}$.

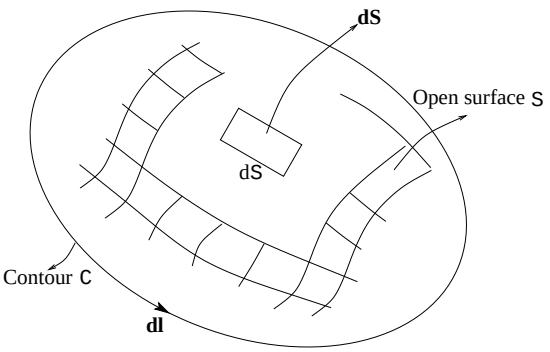


FIGURE 2.16: Defining divergence and gradient using element of volume $\Delta\mathtt{V}$ with surface $\mathtt{S}$.

When Gauss' law is applied to the volume $\Delta\mathtt{V}$, where $\Delta\mathtt{V}$ is small but finite volume, we get

$$\frac{1}{\Delta\mathtt{V}}\oint_{\mathtt{S}} \mathbf{D}\cdot d\mathbf{S} = \frac{1}{\Delta\mathtt{V}}\int_{\mathtt{V}} \rho d\mathtt{V} \tag{2.177}$$

© 2015 by Taylor & Francis Group, LLC

In Figure 2.16, the element of volume $\Delta \mathtt{V}$ containing P has the piece-wise smooth surface S with outward pointing unit normal vector $\hat{\mathbf{n}}$. The divergence of the vector field at P is the limiting value of the net outward flux of $\mathbf{A}$ through the surface S divided by the volume $\Delta \mathtt{V}$, as the volume approaches zero about P. In equation (2.177), the limit $\Delta \mathtt{V}$ approaches zero, and by introducing the divergence theorem (see equation B.25; Appendix B.2.2), we find

$$\oint_{\mathtt{S}} \mathbf{D} \cdot d\mathbf{S} = \int_{\mathtt{V}} \boldsymbol{\nabla} \cdot \mathbf{D} d\mathtt{V}$$

thus,

$$\int_{\mathtt{V}} (\boldsymbol{\nabla} \cdot \mathbf{D} - \rho) d\mathtt{V} = 0 \tag{2.178}$$

which is true for any arbitrary small volume, $d\mathtt{V}$.

Therefore,

$$\boldsymbol{\nabla} \cdot \mathbf{D} = \rho \tag{2.179}$$

This differential relation (equation 2.179) for static electric field is derived from Gauss' law in electrostatics, which states that the electric flux due to an electric field through a closed surface is proportional to the charge contained inside it.

2.4.4 Maxwell's Fourth Law

Maxwell's fourth law is derived from Gauss' magnetic law. The magnetic flux Φ_B through a circuit is equal to the integral of the normal component of the magnetic flux density $\mathbf{B}$ over the surface bounded by the circuit (see equation 2.108). Thus, according to Gauss' law for magnetostatics, the magnetic flux in terms of scalar product is written as

$$\oint_{\mathtt{S}} \mathbf{B} \cdot d\mathbf{S} = 0 \tag{2.180}$$

Unlike Gauss' law for electric field, where the net electric flux through the surface varies as to the net electric charge enclosed by the surface, Gauss' law for magnetic field asserts that there can be no net magnetic flux through the surface owing to the nonavailability of net magnetic charge (individual magnetic poles) enclosed by the surface. Using the divergence theorem, we have

$$\int_{\mathtt{V}} \boldsymbol{\nabla} \cdot \mathbf{B} d\mathtt{V} = 0$$

implying that the flux into the volume element is equal to the flux out of the volume, and thus,

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0 \tag{2.181}$$

Equation (2.181) implies that the magnetic flux due to a magnetic field through a closed surface is zero.

© 2015 by Taylor & Francis Group, LLC

2.4.5 Maxwell's Equations—Sinusoidal Fields

Maxwell's equations in derivative form for a region with sources $\mathbf{J}$ and ρ are described by equations (2.160, 2.175, 2.179, and 2.181); the corresponding equations in integral form are presented in Table A4. However, in most applications, the vector electric and magnetic fields have a sinusoidal time variation. It is worthwhile to represent the fields using the complex exponential form, $e^{i\omega t}(= \cos\omega t + i\sin\omega t)$ (phasor notation), in which ω is the angular frequency. Thus $\mathbf{E} = \mathbf{E}_0 e^{i\omega t}$ and $\mathbf{H} = \mathbf{H}_0 e^{i\omega t}$. The instantaneous value of the field is given by the imaginary part of $\mathbf{E}_0 e^{i\omega t}$, that is,

$$\mathbf{E} = \mathbf{E}_0 \Im e^{i\omega t} = \mathbf{E}_0 \sin\omega t$$

When the sinusoidal fields are expressed in complex exponential form, all time derivatives can be replaced by $i\omega$, i.e., for the electric vector field $\mathbf{H} = \mathbf{H}_0 e^{i\omega t}$, we get

$$\frac{\partial}{\partial t}\mathbf{E} = \frac{\partial}{\partial t}\mathbf{E}_0 e^{i\omega t} = i\omega \mathbf{E}_0 e^{i\omega t} = i\omega \mathbf{E}$$

In a region with sources ρ and $\mathbf{J}$, Maxwell's equations for sinusoidal time-varying fields reduce to

$$\boldsymbol{\nabla}\times\mathbf{E} = -i\omega\mu\mathbf{H} \tag{2.182}$$
$$\boldsymbol{\nabla}\times\mathbf{H} = \sigma\mathbf{E} + i\omega\epsilon\mathbf{E} \tag{2.183}$$
$$\boldsymbol{\nabla}\cdot\mathbf{E} = \rho/\epsilon \tag{2.184}$$
$$\boldsymbol{\nabla}\cdot\mathbf{B} = 0 \tag{2.185}$$
$$\boldsymbol{\nabla}\cdot\mathbf{J} = -i\omega\rho \tag{2.186}$$

It is assumed that the convection current density is zero in equation (2.165); i.e., there exits the conduction current density (see equation 2.164). In a source-free region of space, in which $\sigma = 0$, and $\rho = 0$, the total current density turns out to be zero, i.e., $\mathbf{J} = 0$. Therefore, Maxwell's equations can be cast into the form of wave equation:

$$\boldsymbol{\nabla}\times\mathbf{E} = -\mu_0\frac{\partial\mathbf{H}}{\partial t} = -\frac{\partial\mathbf{B}}{\partial t} \tag{2.187}$$
$$\boldsymbol{\nabla}\times\mathbf{H} = \epsilon_0\frac{\partial\mathbf{E}}{\partial t} = \frac{\partial\mathbf{D}}{\partial t} \tag{2.188}$$
$$\boldsymbol{\nabla}\cdot\mathbf{D} = \boldsymbol{\nabla}\cdot\mathbf{E} = 0 \tag{2.189}$$
$$\boldsymbol{\nabla}\cdot\mathbf{B} = 0 \tag{2.190}$$

In the source-free region is free space $\epsilon = \epsilon_0$ and $\mu = \mu_0$. Maxwell's equations for the sinusoidal time-varying fields are

$$\boldsymbol{\nabla}\times\mathbf{E} = -i\omega\mu_0\mathbf{H} \tag{2.191}$$
$$\boldsymbol{\nabla}\times\mathbf{H} = i\omega\epsilon_0\mathbf{E} \tag{2.192}$$
$$\boldsymbol{\nabla}\cdot\mathbf{E} = 0 \tag{2.193}$$
$$\boldsymbol{\nabla}\cdot\mathbf{B} = 0 \tag{2.194}$$

© 2015 by Taylor & Francis Group, LLC

2.4.6 Continuity Equation of Charge

As stated in Section 2.4.2.3, the Ampère−Maxwell equation led to the continuity equation, so by taking divergence on both sides of equation (2.174), we find

$$\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{H}(\mathbf{r},t)) = \boldsymbol{\nabla}\cdot\mathbf{J}(\mathbf{r},t) + \boldsymbol{\nabla}\cdot\frac{\partial\mathbf{D}(\mathbf{r},t)}{\partial t} \tag{2.195}$$

On applying the vector equation (B.21; see Appendix B.2.2), equation (2.195) translates into

$$\boldsymbol{\nabla}\cdot\mathbf{J}(\mathbf{r},t) = -\boldsymbol{\nabla}\cdot\frac{\partial\mathbf{D}(\mathbf{r},t)}{\partial t} \tag{2.196}$$

By substituting equation (2.179) into equation (2.196), the following relationship emerges:

$$\boldsymbol{\nabla}\cdot\frac{\partial\mathbf{D}(\mathbf{r},t)}{\partial t} = \frac{\partial\rho(\mathbf{r},t)}{\partial t} \tag{2.197}$$

As stated in Section 2.4.2.1, the current density $\mathbf{J}$ is associated with a charge density ρ moving with a velocity. The charge and current density are related by the continuity equation. On replacing the value of $\boldsymbol{\nabla}\cdot\partial\mathbf{D}/\partial t$ from equation (2.196) in equation (2.197), we obtain

$$\boldsymbol{\nabla}\cdot\mathbf{J}(\mathbf{r},t) + \frac{\partial\rho(\mathbf{r},t)}{\partial t} = 0 \tag{2.198}$$

To note, if the volume shrinks down around any point (x, y, z), equation (2.168) turns out to be the same as equation (2.198), which expresses the fact that the net flow of current out of a volume is equal to the negative time rate of change of the volume charge density. This is the equation of continuity, which expresses that the charge is conserved in the neighborhood of any point. By integrating this equation with the help of Gauss' theorem, we find

$$\frac{d}{dt}\int_{\mathtt{V}}\rho d\mathtt{V} + \int_{\mathtt{S}}\mathbf{J}\cdot\hat{\mathbf{n}}d\mathtt{S} = 0 \tag{2.199}$$

The charged particle is a small body with a charge density ρ and the total charge given by equation (2.44) contained within the domain can increase due to flow of electric current, i. Equation (2.98) may written in the form of

$$i = \int_{\mathtt{S}}\mathbf{J}\cdot\hat{\mathbf{n}}d\mathtt{S} \tag{2.200}$$

It is important to note that all the quantities that are figured in Maxwell's equations, as well as in the equation of continuity, are evaluated in the rest frame of the observer and all surfaces and volumes are held fixed in that frame.

© 2015 by Taylor & Francis Group, LLC

2.4.7 Boundary Conditions

Equations (2.34, 2.99, and 2.107) describe the behavior of substances under the influence of the field. These relations are known as material equations. In free space, the relations between the vectors $\mathbf{E}, \mathbf{B}, \mathbf{D}$, and $\mathbf{H}$ in a material are derived from equations (2.34 and 2.107)

$$\begin{aligned} \mathbf{D}(\mathbf{r},t) &= \epsilon \mathbf{E}(\mathbf{r},t) = \epsilon_r \epsilon_0 \mathbf{E}(\mathbf{r},t) \\ \mathbf{H}(\mathbf{r},t) &= \frac{1}{\mu}\mathbf{B}(\mathbf{r},t) = \frac{1}{\mu_r \mu_0}\mathbf{B}(\mathbf{r},t) \end{aligned} \tag{2.201}$$

where ϵ is the permittivity of the medium in which the electric field acts, μ the permeability of the medium in which the magnetic field acts, and $\epsilon_r (= \epsilon/\epsilon_0)$ and $\mu_r (= \mu/\mu_0)$ are, respectively, the relative permittivity and the relative permeability of the medium.

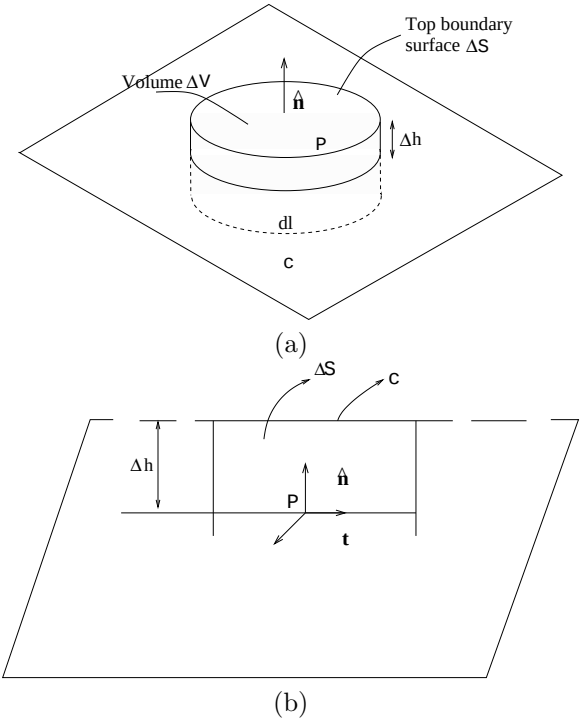


FIGURE 2.17: (a) Boundary conditions for the normal components of the electromagnetic field and (b) the integration path in the boundary surface.

The additional fields are present in matter giving rise to the notation

$$\mathbf{E}_m = \mathbf{E} + \frac{\mathbf{P}}{\epsilon_0} \tag{2.202}$$

$$\mathbf{B}_m = \mathbf{B} + \mu_0 \mathbf{M} \tag{2.203}$$

© 2015 by Taylor & Francis Group, LLC

where $\mathbf{E}_m$ is the electric field corresponding to the dielectric displacement in $\mathtt{V}.m^{-1}$, $\mathbf{B}_m$ the magnetic field in the presence of medium, $\mathbf{P}$ the polarization, $\mathbf{M}$ the magnetization, and μ_0 the permeability in vacuum.

Let us assume that both ϵ and μ in equation (2.201) are independent of position $\mathbf{r}$ and time t, and that $\epsilon_r \geq 1, \mu_r \geq 1$. The field vectors can be determined in regions of space (Figure 2.17a) where both ϵ and μ are continuous functions of space from the set of Maxwell's equations, as well as from the material equations. As stated in Section 2.4.4, the flux into the volume element is equal to the flux out of the volume. For a flat volume whose faces can be neglected; therefore, from equation (2.180), we get

$$\oint_{\mathtt{S}} \mathbf{B}\cdot\hat{\mathbf{n}} d\mathtt{S} = 0 \tag{2.204}$$

Similarly, Maxwell's third equation (2.179) may also be used. With boundary conditions at the interface between two different media, i.e., when the physical properties of the medium are discontinuous, the electromagnetic fields within a bounded region Ω are given by

$$\hat{\mathbf{n}}\cdot(\mathbf{B}_2 - \mathbf{B}_1) = 0 \tag{2.205}$$

$$\hat{\mathbf{n}}\cdot(\mathbf{D}_2 - \mathbf{D}_1) = \rho \tag{2.206}$$

in which $\hat{\mathbf{n}}$ is the unit vector normal to the surface of discontinuity directed from medium 1 to medium 2 and Λ the surface density of bound charge.

Equations (2.205, 2.206) may be written as

$$\mathbf{B}_{2n} - \mathbf{B}_{1n} = 0 \tag{2.207}$$

$$\mathbf{D}_{2n} - \mathbf{D}_{1n} = \rho \tag{2.208}$$

where $\mathbf{B}_n = \hat{\mathbf{n}}\cdot\mathbf{B}$ and the subscript n signifies the component normal to the boundary surface.

Equations (2.207) and (2.208) reveal the boundary conditions for the normal components of $\mathbf{B}$ and $\mathbf{D}$, respectively. The normal component of magnetic induction is continuous, while the normal component of electric displacement changes across the boundary as a result of surface charges. From the Ampère−Maxwell equation (2.174), the condition for $\mathbf{H}$ can be derived. Let us choose the integration path in a way that the unit vector is tangential to the interface between the media (Figure 2.17b). The integral form of the equation after applying Stokes formula yields

$$\mathbf{t}\times(\mathbf{H}_2 - \mathbf{H}_1) = \mathbf{J}_s \tag{2.209}$$

where $\mathbf{t}$ signifies the unit vector tangential to the interface between the media, and $\mathbf{J}_s$ the surface density of current tangential to the interface, locally perpendicular to both $\mathbf{t}$ and $\hat{\mathbf{n}}$.

Similarly, for a static case, a corresponding equation for the tangent component of the electric field as given in equation (2.84) is written as

$$\mathbf{t}\times(\mathbf{E}_2 - \mathbf{E}_1) = 0 \tag{2.210}$$

© 2015 by Taylor & Francis Group, LLC

Equations (2.209 and 2.210) demonstrate respectively that the tangential components of the electric field vector are continuous across the boundary and the tangential component of the magnetic vector changes across the boundary as a result of a surface current density. Since $\mathbf{B} = \mu\mathbf{H}$, from equation (2.205), we have

$$\mu_1(\mathbf{H}_1 \cdot \hat{\mathbf{n}}) = \mu_2(\mathbf{H}_2 \cdot \hat{\mathbf{n}}) \tag{2.211}$$

and for the normal component,

$$(\mathbf{H}_1)_n = \frac{\mu_2}{\mu_1}(\mathbf{H}_2)_n \tag{2.212}$$

In the case of the equation of continuity for electric charge (equation 2.198), the boundary condition is given by

$$\hat{\mathbf{n}} \cdot (\mathbf{J}_2 - \mathbf{J}_1) + \boldsymbol{\nabla}_s \cdot \mathbf{J}_s = -\frac{\partial \rho_s}{\partial t} \tag{2.213}$$

This is the surface equation of continuity for electric charge; it is a statement of conservation of charge at a point on the surface.

2.5 Energy Flux of Electrodynamics

Concepts such as energy, linear, and angular momentum may be associated with the electromagnetic field through equation (2.131; see Section 2.2.5). The angular momentum is defined as the product of moment of inertia and angular velocity of a body revolving about an axis. In classical mechanics, a particle of mass m moving with a velocity $\mathbf{v}$ at position $\mathbf{r}$ in an inertial reference frame has linear momentum $\mathbf{p}$ ([125], [131]):

$$\mathbf{p} = m\frac{d\mathbf{r}}{dt} = m\mathbf{v} \tag{2.214}$$

According to the Lorentz force law (see equation 2.131), the equation of motion of a free particle of charge q and mass m moving in electric and magnetic fields is

$$m\frac{d\mathbf{v}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{2.215}$$

The total force applied to the particle is attributed to the Newton second law:

$$\begin{aligned} \mathbf{F} &= \frac{d\mathbf{p}}{dt} = m\frac{d\mathbf{v}}{dt} \\ &= m\frac{d^2\mathbf{r}}{dt^2} = m\mathbf{a} \end{aligned} \tag{2.216}$$

in which $\mathbf{a} = \frac{q}{m}\mathbf{E}$ indicates the acceleration of the particle, $\mathrm{m.s}^{-2}$.

© 2015 by Taylor & Francis Group, LLC

The force $\mathbf{F}$ on the particle is equal to the charge of a particle that is placed in an uniform electric field, i.e., $\mathbf{F} = q\mathbf{E}$. The force is in the same direction as the field if the charge is positive, and the force becomes opposite to the field if the charge is negative. If the particle is at rest and the field is applied, the particle is accelerated uniformly in the direction of the field. The equation of motion (equation 2.216) was verified in an experiment carried out by J. J. Thompson (1856−1940) using a cathode ray tube[42].

The work done by the applied force on the particle when it moves through the displacement $\Delta\mathbf{r}$ is defined as

$$\Delta W = \mathbf{F}\cdot\Delta\mathbf{r} \tag{2.217}$$

The rate at which the work is done is the power $\mathcal{P}$:

$$\begin{aligned}\mathcal{P} &= \lim_{\Delta t\to 0}\left(\frac{\Delta W}{\Delta t}\right)\\ &= \lim_{\Delta t\to 0}\left(\mathbf{F}\cdot\frac{\Delta\mathbf{r}}{\Delta t}\right) = \mathbf{F}\cdot\mathbf{v}\end{aligned} \tag{2.218}$$

It follows from the Lorentz force law that the power input to a particle moving in electric and magnetic fields is

$$\mathcal{P} = q\mathbf{v}\cdot\mathbf{E} \tag{2.219}$$

The energy in the case of a continuous charge configuration $\rho(\mathbf{r})$ is expressed as

$$W = \frac{1}{2}\iint\frac{\rho(\mathbf{r}')\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}d\mathtt{V}d\mathtt{V}' = \frac{1}{2}\int\varphi(\mathbf{r})\rho(\mathbf{r})d\mathtt{V} \tag{2.220}$$

where $\varphi(\mathbf{r})$ is the potential of a charge distribution (see equation 2.69).

In equation (2.220), the integration extends over the point $\mathbf{r} = \mathbf{r}'$, so that the said equation contains self-energy parts which become infinitely large for point charges. The amount of electrostatic energy stored in an electric field in a region of space is expressed as

$$\begin{aligned}W &= \frac{1}{2}\int\varphi(\mathbf{r})\rho(\mathbf{r})d\mathtt{V} = \frac{1}{2}\int\left[\boldsymbol{\nabla}\cdot\mathbf{E}(\mathbf{r})\right]\varphi(\mathbf{r})d\mathtt{V}\\ &= -\frac{1}{2}\int E(\mathbf{r})\cdot\boldsymbol{\nabla}\varphi(\mathbf{r})d\mathtt{V} = \frac{1}{2}\int E^2(\mathbf{r})d\mathtt{V}\end{aligned} \tag{2.221}$$

The integrand represents the energy density associated with the electric field, that is,

$$w_e = \frac{1}{2}\mathbf{E}^2 \tag{2.222}$$

The power can be determined in terms of the kinetic energy (KE) of the particle $\mathcal{K}$ by invoking equation (2.216):

$$\begin{aligned}\mathcal{P} &= \mathbf{F}\cdot\mathbf{v} = m\frac{d\mathbf{v}}{dt}\cdot\mathbf{v}\\ &= \frac{d}{dt}\left(\frac{1}{2}m|\mathbf{v}|^2\right) = \frac{d\mathcal{K}}{dt}\end{aligned} \tag{2.223}$$

© 2015 by Taylor & Francis Group, LLC

Thus, the rate at which work is done by the applied force − the power − is equal to the rate of increase in KE of the particle.

Let us examine a charged particle moving in an electromagnetic field, $\mathbf{E}, \mathbf{B}$, in free space. The mechanical force of electromagnetic origin acting on the charge and current for a volume $\mathtt{V}$ of free space at rest containing charge ρ and current density $\mathbf{J}$ can also be given by the Lorentz law. Equation (2.131) can be modified to

$$\mathbf{F} = \int_{\mathtt{V}} (\rho\mathbf{E} + \mathbf{J}\times\mathbf{B})\, d\mathtt{V} \tag{2.224}$$

where $\mathbf{J}(= \rho\mathbf{v})$ is the current density and $\mathbf{v}$ the velocity of the particle moving the current density within the particle. From equation (2.131), we derive the power $\mathcal{P}$ as

$$\begin{aligned} \mathcal{P} &= \mathbf{v}\cdot \int_{\mathtt{V}} (\rho\mathbf{E} + \rho\mathbf{v}\times\mathbf{B})\, d\mathtt{V} \\ &= \int_{\mathtt{V}} [\rho\mathbf{v}\cdot\mathbf{E} + \mathbf{v}\cdot(\rho\mathbf{v}\times\mathbf{B})]\, d\mathtt{V} \end{aligned} \tag{2.225}$$

Since the velocity is same at all points in the particle, $\mathbf{v}$ is moved under the integral sign. According to the vector relation, divergence of a cross product is zero (see equation B.24; Appendix B.2.2); therefore, the magnetic field does no work on the charged particle. Thus, equation (2.225) is written as

$$\mathcal{P} = \int_{\mathtt{V}} \mathbf{E}\cdot\mathbf{J} d\mathtt{V} = \frac{d\mathcal{K}}{dt} \tag{2.226}$$

Equation (2.226) expresses the rate at which energy is exchanged between the electromagnetic field and the mechanical motion of the charged particle. When $\mathcal{P}$ is positive, the field supplies energy to the mechanical motion of the particle, and in the case of negative $\mathcal{P}$, the mechanical motion of the particle supplies motion to the field.

2.5.1 Poynting Vector

The energy conservation law of the electromagnetic field was evolved by Poynting[43] from Maxwell's equations (2.175 and 2.160), which results, respectively, in

$$\mathbf{E}\cdot(\boldsymbol{\nabla}\times\mathbf{H}) = \mathbf{E}\cdot\mathbf{J} + \mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t} \tag{2.227}$$

$$\mathbf{H}\cdot(\boldsymbol{\nabla}\times\mathbf{E}) = -\mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t} \tag{2.228}$$

By subtracting equation (2.228) from equation (2.227), we get

$$\mathbf{E}\cdot(\boldsymbol{\nabla}\times\mathbf{H}) - \mathbf{H}\cdot(\boldsymbol{\nabla}\times\mathbf{E}) = \mathbf{E}\cdot\mathbf{J} + \mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t} + \mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t} \tag{2.229}$$

© 2015 by Taylor & Francis Group, LLC

where the first term on the RHS of equation (2.229), i.e., $\mathbf{E}\cdot\mathbf{J}$, represents the work done by the field on the electric current density; the second term, $\mathbf{E}\cdot\frac{\partial \mathbf{D}}{\partial t}$, the density of reactive power driving the build-up of electric field; and the last term, $\mathbf{H}\cdot\frac{\partial \mathbf{B}}{\partial t}$, the density of reactive power driving the build-up of magnetic field.

By using the vector relation given in equation (B.34; Appendix B.2.4), the LHS quantity of equation (2.229) can be written as

$$\mathbf{E}\cdot(\boldsymbol{\nabla}\times\mathbf{H}) - \mathbf{H}\cdot(\boldsymbol{\nabla}\times\mathbf{E}) = -\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{H}) \tag{2.230}$$

which turns out to be

$$\mathbf{E}\cdot\mathbf{J} + \mathbf{E}\cdot\frac{\partial \mathbf{D}}{\partial t} + \mathbf{H}\cdot\frac{\partial \mathbf{B}}{\partial t} + \boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{H}) = 0 \tag{2.231}$$

On applying Gauss' divergence theorem (see equation B.25; Appendix B.2.2), followed by integrating all through an arbitrary volume, equation (2.231) is deduced as

$$\int_{\mathtt{V}} \mathbf{E}\cdot\mathbf{J}\, d\mathtt{V} + \int_{\mathtt{V}} \left(\mathbf{E}\cdot\frac{\partial \mathbf{D}}{\partial t} + \mathbf{H}\cdot\frac{\partial \mathbf{B}}{\partial t}\right) d\mathtt{V} + \oint_{\mathtt{S}} (\mathbf{E}\times\mathbf{H})\cdot d\mathbf{S} = 0 \tag{2.232}$$

By defining

$$\begin{aligned} \mathbf{S}(\mathbf{r},t) &= \mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t) \\ &= \frac{1}{\mu_0}\mathbf{E}(\mathbf{r},t)\times\mathbf{B}(\mathbf{r},t) \end{aligned} \tag{2.233}$$

we derive the energy flux carried by the electromagnetic field. Here $\mathbf{H} = \mathbf{B}/\mu_0$ (see equation 2.107) and $\mathbf{r}(= x, y, z)$ is the 3D position vector. The term $\mathbf{S}$ is called the energy flux density in the direction of propagation. It is known as the Poynting vector, or power surface density. It is interpreted as the time derivative of the sum of the electrostatic and magnetic energy densities. The Poynting vector $\mathbf{S}$ has the units of energy per unit area per unit time ($\mathrm{J.m^{-2}.s^{-1}}$) or power per unit area ($\mathrm{W.m^{-2}}$). Its magnitude $|\mathbf{S}|$ is equal to the rate of flow per unit area element perpendicular to $\hat{\mathbf{S}}$. Since the fields $\mathbf{E}$ and $\mathbf{B}$ are perpendicular, we may verify that the magnitude of $\mathbf{S}$ is

$$|\ \mathbf{S}\ | = \frac{EB}{\mu_0} \tag{2.234}$$

2.5.2 Energy Conservation Law of the Electromagnetic Field

Two ideas, namely, wave equation and energy were borrowed from mechanics in order to arrive at the mathematical description of the interference fringes that are observed from the Young experiment (see Section 1.6.2). The solution of the former was inadequate to discuss such fringes. Describing the

© 2015 by Taylor & Francis Group, LLC

optical field in terms of energy, known as intensity in optics, led to results in agreement with the fringes with respect to their intensity and spacing. As stated in Section 1.5, the intensity is formed by taking time averages of the quadratic field components over an interval much greater than the time period $T = 2\pi/\omega$. Electromagnetic theory interprets the light intensity as the energy flux of the field.

Thus far the expression we obtained in the preceding Section 2.5.1 is for the energy associated with the motion of a charged particle. Electromagnetic theory interprets the light intensity as the energy flux of the field. In what follows, an expression for the energy that applies to the general volume distribution of charge ρ and current $\mathbf{J}$ is derived. Let us rewrite equation (2.232) in the form

$$\int_{\mathtt{V}} \mathbf{E}\cdot\mathbf{J}\, d\mathtt{V} + \int_{\mathtt{V}} \left(\mathbf{E}\cdot\frac{\partial \mathbf{D}}{\partial t} + \mathbf{H}\cdot\frac{\partial \mathbf{B}}{\partial t} \right) d\mathtt{V} + \oint_{\mathtt{S}} \mathbf{S}\cdot d\mathbf{S} = 0 \tag{2.235}$$

When integrated over a closed surface the normal component of the Poynting vector provides the total outward flow of the energy per unit time. This relation is known as Poynting theorem. The power carried away from a volume bounded by a surface $\mathtt{S}$ by the electromagnetic field is given by the term $\oint_{\mathtt{S}} \mathbf{S}\cdot d\mathbf{S}$. On using material equations (2.34 and 2.107), the second term of the Poynting theorem (equation 2.233) can be simplified. For the electric term, we have

$$\mathbf{E}\cdot\frac{\partial \mathbf{D}}{\partial t} = \mathbf{E}\cdot\frac{\partial}{\partial t}\left(\epsilon_0 \mathbf{E}\right) = \frac{1}{2}\frac{\partial}{\partial t}\left(\epsilon_0 \mathbf{E}^2\right) = \frac{1}{2}\frac{\partial}{\partial t}\left(\mathbf{E}\cdot\mathbf{D}\right) \tag{2.236}$$

Similarly, for the magnetic term one may derive

$$\mathbf{H}\cdot\frac{\partial \mathbf{B}}{\partial t} = \frac{1}{2}\frac{\partial}{\partial t}\left(\mu_0 \mathbf{H}^2\right) = \frac{1}{2}\frac{\partial}{\partial t}\left(\mathbf{H}\cdot\mathbf{B}\right) \tag{2.237}$$

Thus, the second term of equation (2.235) is recast as

$$\int_{\mathtt{V}} \left(\mathbf{E}\cdot\frac{\partial \mathbf{D}}{\partial t} + \mathbf{H}\cdot\frac{\partial \mathbf{B}}{\partial t} \right) d\mathtt{V} = \frac{1}{2}\frac{\partial}{\partial t}\int_{\mathtt{V}} \left(\mathbf{E}\cdot\mathbf{D} + \mathbf{H}\cdot\mathbf{B}\right) d\mathtt{V} \tag{2.238}$$

For an electrostatic field in a simple material, the energy stored in the electric field and, for a magnetostatic field in a simple material, the stored energy in the magnetic field are respectively given by

$$w_e = \frac{1}{2}\mathbf{E}\cdot\mathbf{D} = \frac{1}{2}\epsilon_0 \mathbf{E}^2 \tag{2.239}$$

$$w_m = \frac{1}{2}\mathbf{H}\cdot\mathbf{B} = \frac{1}{2}\mu_0 \mathbf{H}^2 \tag{2.240}$$

where w_e and w_m are the energy densities associated with the electric and magnetic fields, respectively.

© 2015 by Taylor & Francis Group, LLC

From equations (2.238−2.240), we rewrite equation (2.231) as

$$\mathbf{E}\cdot\mathbf{J} + \boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{H}) = \frac{\partial}{\partial t}(w_e + w_m) \tag{2.241}$$

Equation (2.241) describes the transfer of energy during a decrease of the total energy density of the electromagnetic field in time. Let us calculate the divergence of Stokes vector (second term on the LHS of equation 2.241) for any electromagnetic field in free space. On using equation (B.33; Appendix B.2.4), we may write

$$\begin{aligned}\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{H}) &= -\mathbf{E}\cdot(\boldsymbol{\nabla}\times\mathbf{H}) + \mathbf{H}\cdot(\boldsymbol{\nabla}\times\mathbf{E}) \\ &= -\mathbf{E}\frac{\partial\mathbf{D}}{\partial t} - \mathbf{H}\frac{\partial\mathbf{B}}{\partial t} \\ &= -\frac{\partial}{\partial t}\left[\frac{1}{2}\epsilon_0\mathbf{E}^2 + \frac{1}{2}\mu_0\mathbf{H}^2\right]\end{aligned} \tag{2.242}$$

Integrating over a volume $\mathtt{V}$, bounded by a surface $\mathtt{S}$, and using the divergence theorem (see Appendix B.2.2), we have

$$\int_{\mathtt{S}} (\mathbf{E}\times\mathbf{H})\cdot d\mathbf{S} = -\frac{\partial}{\partial t}\int_{\mathtt{V}}\left[\frac{1}{2}\epsilon_0\mathbf{E}^2 + \frac{1}{2}\mu_0\mathbf{H}^2\right] d\mathtt{V} \tag{2.243}$$

The integral on the RHS, the sum of the electric energy, w_e, and magnetic energy, w_m, is the energy lost per unit time by the volume, $\mathtt{V}$, and the LHS must be the total outward flux of energy in watts over the surface $\mathtt{S}$ bounding the volume $\mathtt{V}$. The Poynting theorem (equation 2.235) takes the form

$$\begin{aligned}-\frac{dW}{dt} &= \frac{d}{dt}\int_{\mathtt{V}}(w_e + w_m)d\mathtt{V} \\ &= \int_{\mathtt{V}}\mathbf{E}\cdot\mathbf{J}d\mathtt{V} + \oint_{\mathtt{S}}\mathbf{S}\cdot d\mathbf{S}\end{aligned} \tag{2.244}$$

in which $W = \int_{\mathtt{V}}(w_e + w_m)d\mathtt{V}$ is the total energy.

Equation (2.244) represents the energy conservation law of electrodynamics. The term dW/dt is interpreted as the time rate of change of the total energy contained within the volume. Let us recall the Lorentz law (equation 2.131) and assume that all the charges q_k are displayed by $\delta\mathbf{x}_k$ (where $k = 1, 2, 3, \cdots$) in time, δt; therefore, the total work done is given by

$$\begin{aligned}\delta A &= \sum_k q_k\left[\mathbf{E}_k + \mathbf{v}_k\times\mathbf{B}\right]\cdot\delta\mathbf{x}_k \\ &= \sum_k q_k\mathbf{E}_k\cdot\delta\mathbf{x}_k = \sum_k q_k\mathbf{E}_k\cdot\mathbf{v}_k\delta t\end{aligned} \tag{2.245}$$

where $\delta\mathbf{x}_k = \mathbf{v}_k\delta t$.

© 2015 by Taylor & Francis Group, LLC

On introducing the total charge density ρ, we obtain

$$\frac{\delta A}{\delta t} = \int_{\mathtt{V}} \rho \mathbf{v} \cdot \mathbf{E} d\mathtt{V} \tag{2.246}$$

As stated earlier in Section 2.4.2.1, the current density $\mathbf{J}$ (see equation 2.99) comprises two parts, $\mathbf{J}_c$ and $\mathbf{J}_v$; hence, for an isothermal conductor, the energy is irreversibly transferred to a heat reservoir as the Joule heat (James Brescott Joule, 1818−1889), so we write

$$\mathcal{Q} = \int_{\mathtt{V}} \mathbf{E} \cdot \mathbf{J}_c d\mathtt{V} = \int_{\mathtt{V}} \sigma \mathbf{E}^2 d\mathtt{V} \tag{2.247}$$

Here, $\mathcal{Q}$ represents resistive dissipation of energy called the Joule heat in a conductor ($\sigma \neq 0$). When the motion of the charge is instantaneously supplying energy to the electromagnetic field throughout the volume, the volume density of current due to the motion of the charge $\mathbf{J}_v$ is given by

$$\frac{\delta A}{\delta t} = \int_{\mathtt{V}} \mathbf{E} \cdot \mathbf{J}_v d\mathtt{V} \tag{2.248}$$

From equations (2.165 and 2.246), we obtain

$$\int_{\mathtt{V}} \mathbf{J} \cdot \mathbf{E} d\mathtt{V} = \mathcal{Q} + \int_{\mathtt{V}} \mathbf{J}_v \cdot \mathbf{E} d\mathtt{V} = \mathcal{Q} + \frac{\delta A}{\delta t} \tag{2.249}$$

Thus, equation (2.244) translates into

$$\frac{dW}{dt} = -\mathcal{Q} - \frac{\delta A}{\delta t} - \oint_{\mathtt{S}} \mathbf{S} \cdot d\mathbf{S} \tag{2.250}$$

where $\delta A/\delta t$ is the rate at which electromagnetic energy is being stored.

The interpretation of such a relation as a statement of conservation of energy within the volume stands. For a nonconductor, when $\mathcal{Q}$ is zero, dW/dt turns out to be zero; therefore, equation (2.250) becomes

$$\oint_{\mathtt{S}} \mathbf{S} \cdot d\mathbf{S} = -\frac{\delta A}{\delta t} \tag{2.251}$$

In a nonconducting medium ($\sigma = 0$) where no mechanical work is done ($A = 0$), the energy law may be written in the hydrodynamical continuity equation for noncompressible fluids:

$$\frac{\partial w}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{S} = 0 \tag{2.252}$$

The physical meaning of equation (2.252) is that the decrease in the time rate of change of electromagnetic energy within a volume is equal to the flow of energy out of the volume.

© 2015 by Taylor & Francis Group, LLC

2.6 Electromagnetic Field Equations

Let us recast equation (2.187) as

$$\boldsymbol{\nabla}\times\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} \tag{2.253}$$

The curl (see Appendix B.2.4) of equation (2.253) provides us with

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{E}) = -\mu_0 \frac{\partial}{\partial t}(\boldsymbol{\nabla}\times\mathbf{H}) \tag{2.254}$$

Differentiating both sides of equation (2.188) with respect to time and interchanging differentiation with respect to time and space, we get

$$\boldsymbol{\nabla}\times\frac{\partial \mathbf{H}}{\partial t} = \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{2.255}$$

The wave equations for both electric and magnetic fields can be derived as follows:

1. Wave equation for **E**: Substituting equation (2.255) into equation (2.254), the following relationship emerges

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{E}) = -\mu_0\epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{2.256}$$

By using the vector triple product identity (see equation B.33; Appendix B.2.4), we may write

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{E}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{E}) - \boldsymbol{\nabla}^2\mathbf{E} \tag{2.257}$$

In vacuum, $\boldsymbol{\nabla}\cdot\mathbf{E} = 0$, so

$$\begin{aligned} \boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{E}) &= -\boldsymbol{\nabla}^2\mathbf{E} = -\mu_0\frac{\partial}{\partial t}(\boldsymbol{\nabla}\times\mathbf{H}) \\ &= -\mu_0\frac{\partial}{\partial t}\left(\frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}\right) = -\boldsymbol{\nabla}^2\mathbf{E} \end{aligned} \tag{2.258}$$

Here $\mathbf{J} = 0$ for source free region. Invoking equation (2.254), equation (2.258) translates into

$$\boldsymbol{\nabla}^2\mathbf{E} = \mu_0\epsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2}$$

or

$$\left(\boldsymbol{\nabla}^2 - \mu_0\epsilon_0\frac{\partial^2}{\partial t^2}\right)\mathbf{E} = 0 \tag{2.259}$$

© 2015 by Taylor & Francis Group, LLC

The vector equation (2.259) consists of three separate partial differential equations:

$$\frac{\partial^2 E_x}{\partial x^2} = \mu_0\epsilon_0\frac{\partial^2 E_x}{\partial t^2}; \quad \frac{\partial^2 E_y}{\partial y^2} = \mu_0\epsilon_0\frac{\partial^2 E_y}{\partial t^2}; \quad \frac{\partial^2 E_z}{\partial z^2} = \mu_0\epsilon_0\frac{\partial^2 E_z}{\partial t^2} \tag{2.260}$$

2. Wave equation for **H**: Following equation (2.257),

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{H}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{H}) - \boldsymbol{\nabla}^2\mathbf{H}$$

the wave equation for the magnetic field, **H**, is obtained as

$$\begin{aligned}\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{H}) &= \boldsymbol{\nabla}\times\left[\left(\mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t} + \mu_0\mathbf{J}\right)\right] \\ &= \mu_0\epsilon_0\frac{\partial}{\partial t}(\boldsymbol{\nabla}\times\mathbf{E}) \\ &= -\mu_0\epsilon_0\frac{\partial^2\mathbf{H}}{\partial t^2}\end{aligned} \tag{2.261}$$

The wave equation is

$$\boldsymbol{\nabla}^2\mathbf{H} = \mu_0\epsilon_0\frac{\partial^2\mathbf{H}}{\partial t^2}$$

or

$$\left(\boldsymbol{\nabla}^2 - \mu_0\epsilon_0\frac{\partial^2}{\partial t^2}\right)\mathbf{H} = 0 \tag{2.262}$$

There are four such families of equations for the four field vectors, $\mathbf{E}, \mathbf{D}, \mathbf{H}$, and $\mathbf{B}$. The above expressions (equations 2.259 and 2.262), known as the electromagnetic wave equations, satisfy the classical wave equation with velocity of propagation.

3. Velocity of light: The permittivity constant, ϵ_0, and the permeability constant, μ_0, in a vacuum are related to the speed of light, c, by

$$c = \frac{1}{\sqrt{\mu_0\epsilon_0}} = 2.99,79\times10^8 \text{ m.s}^{-1} \tag{2.263}$$

therefore, we may express the wave equation (2.262) as

$$\left(\boldsymbol{\nabla}^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{E} = 0 \tag{2.264}$$

4. Time-independent wave equations: Let us write

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0(\mathbf{r})e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)} \tag{2.265}$$

$$\mathbf{B}(\mathbf{r},t) = \mathbf{B}_0(\mathbf{r})e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)} \tag{2.266}$$

© 2015 by Taylor & Francis Group, LLC

where $\mathbf{E}_0$ and $\mathbf{B}_0$ are constant vectors, and $\mathbf{r}(= x, y, z)$ is the 3D position vector.

Therefore, the respective time-independent wave equations are derived as

$$\boldsymbol{\nabla}^2\mathbf{E}(\mathbf{r},t) + \frac{\omega^2}{c^2}\mathbf{E}_0(\mathbf{r}) = 0 \tag{2.267}$$

$$\boldsymbol{\nabla}^2\mathbf{B}(\mathbf{r},t) + \frac{\omega^2}{c^2}\mathbf{B}_0(\mathbf{r}) = 0 \tag{2.268}$$

in which $\omega(= c\kappa)$ is the angular frequency and κ the wavenumber (see Section 1.2).

The time-independent wave equations are given as

$$\boldsymbol{\nabla}^2\mathbf{E}(\mathbf{r},t) + \kappa^2\mathbf{E}(\mathbf{r}) = 0 \tag{2.269}$$

$$\boldsymbol{\nabla}^2\mathbf{B}(\mathbf{r},t) + \kappa^2\mathbf{B}(\mathbf{r}) = 0 \tag{2.270}$$

2.6.1 General Electromagnetic Wave

It is evident from Maxwell's equations that the harmonic variations of the electric and magnetic fields are always perpendicular to each other and to the direction of propagation denoted by $\boldsymbol{\kappa}$ (see Figure 1.5). These variations are described by the harmonic wave equations in the form

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0(\mathbf{r},\omega)e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)} \tag{2.271}$$

$$\mathbf{B}(\mathbf{r},t) = \mathbf{B}_0(\mathbf{r},\omega)e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)} \tag{2.272}$$

in which $\mathbf{E}_0(\mathbf{r},\omega)$ and $\mathbf{B}_0(\mathbf{r},\omega)$ are the amplitudes of the electric and magnetic field vectors, respectively; ω is the angular frequency, and T the period of motion (see Section 1.2), and

$$\boldsymbol{\kappa}\cdot\mathbf{r} = \kappa_x x + \kappa_y y + \kappa_z z \tag{2.273}$$

represents planes in space of constant phase, and $\boldsymbol{\kappa}$ the wave vector defined by equation (1.15).

Both the Cartesian components of the wave travel with the same propagation vector $\boldsymbol{\kappa}$ and frequency ω. The cosinusoidal fields are

$$\mathbf{E}(\mathbf{r},t) = \Re\left[\mathbf{E}_0(\mathbf{r},\omega)e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)}\right] = \mathbf{E}_0(\mathbf{r},\omega)\cos(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t) \tag{2.274}$$

$$\mathbf{B}(\mathbf{r},t) = \Re\left[\mathbf{B}_0(\mathbf{r},\omega)e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)}\right] = \mathbf{B}_0(\mathbf{r},\omega)\cos(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t) \tag{2.275}$$

Let us assume that $\mathbf{E}_0$ is constant, so the divergence of the equation (2.271) is given by

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{E} &= \mathbf{E}_0\cdot\boldsymbol{\nabla}\left(e^{i[\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t]}\right)\\ &= \mathbf{E}_0\cdot(i\boldsymbol{\kappa})e^{i[\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t]} = (i\boldsymbol{\kappa})\cdot\mathbf{E}\end{aligned} \tag{2.276}$$

© 2015 by Taylor & Francis Group, LLC

The curl of the electric field is derived as

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{H} &= \frac{\epsilon_0}{c}\frac{\partial\mathbf{E}}{\partial t} = \frac{\epsilon_0}{c}\mathbf{E}_0\frac{\partial}{\partial t}\left(e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)}\right)\\ &= -\frac{i\omega\epsilon_0}{c}\mathbf{E}_0 e^{i(\boldsymbol{\kappa}\cdot\mathbf{r}-\omega t)} = -\frac{i\omega\epsilon_0}{c}\mathbf{E}\end{aligned} \tag{2.277}$$

By replacing $\boldsymbol{\nabla}\rightarrow i\boldsymbol{\kappa}$ and $\partial/\partial t\rightarrow -i\omega$, we rewrite equation (2.277) as

$$\boldsymbol{\kappa}\times\mathbf{H} = -\frac{\epsilon_0\omega}{c}\mathbf{E} \tag{2.278}$$

Similarly, with another Maxwell's equation (2.160), we derive

$$\boldsymbol{\kappa}\times\mathbf{E} = \frac{\omega\mu_0}{c}\mathbf{H} \tag{2.279}$$

After rearranging equations (2.278 and 2.279), we find

$$\mathbf{E} = -\frac{c}{\epsilon_0\omega}\boldsymbol{\kappa}\times\mathbf{H} = -\frac{1}{\omega}\sqrt{\frac{\mu_0}{\epsilon_0}}\boldsymbol{\kappa}\times\mathbf{H} \tag{2.280}$$

$$\mathbf{H} = \frac{c}{\omega\mu_0}\boldsymbol{\kappa}\times\mathbf{E} = \frac{1}{\omega}\sqrt{\frac{\epsilon_0}{\mu_0}}\boldsymbol{\kappa}\times\mathbf{E} \tag{2.281}$$

Since the divergence of both the electric and magnetic fields is zero, i.e., $\boldsymbol{\nabla}\cdot\mathbf{E} = 0$. and $\boldsymbol{\nabla}\cdot\mathbf{B} = 0$, we derive

$$\boldsymbol{\kappa}\cdot\mathbf{E} = 0 \tag{2.282}$$

$$\boldsymbol{\kappa}\cdot\mathbf{B} = 0 \tag{2.283}$$

indicating that the electric and magnetic field vectors lie in planes normal to the direction of propagation that is represented by the wave vector, $\boldsymbol{\kappa}$. In other words, the electromagnetic waves are transverse waves. Scalar multiplication with $\boldsymbol{\kappa}$ provides us with

$$\mathbf{E}\cdot\boldsymbol{\kappa} = \mathbf{H}\cdot\boldsymbol{\kappa} = 0 \tag{2.284}$$

From equation (2.284) we get

$$\sqrt{\mu_0}|\mathbf{H}| = \sqrt{\epsilon_0}|\mathbf{E}| \tag{2.285}$$

The magnitude of a real vector $|\mathbf{E}|$ for a general time-dependent electromagnetic field $\mathbf{E}(\mathbf{r},t)$ is represented by $\sqrt{\mathbf{E}\cdot\mathbf{E}}$. In Cartesian coordinates the quadratic term, $\mathbf{E}\cdot\mathbf{E}$, is written out as

$$\mathbf{E}\cdot\mathbf{E} = E_xE_x + E_yE_y \tag{2.286}$$

Thus, Maxwell's theory leads to quadratic terms associated with the flow of energy, called the intensity. It is observed from equations (2.276−2.279) that in an electromagnetic wave, the field intensities $\mathbf{E},\mathbf{H}$ and the unit vector in the propagation direction of the wave $\boldsymbol{\kappa}$ form a right-handed orthogonal triad

© 2015 by Taylor & Francis Group, LLC

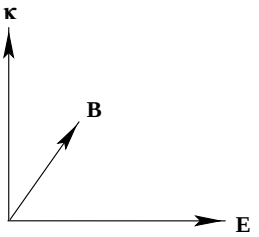


FIGURE 2.18: Orthogonal triad of vectors.

of vectors (see Figure 2.18). To be precise, if an electromagnetic wave travels in the positive x axis, the electric and magnetic fields would oscillate parallel to the y and z axis, respectively.

The energy crossing an element of area in unit time is perpendicular to the direction of propagation. In a cylinder whose axis is parallel to $\mathbf{S}$ with unit cross-sectional area, the amount of energy passing the base of the cylinder in unit time is equal to the energy that is contained in the portion of the cylinder of length v. Therefore, the energy flux is equal to vw, where

$$w = \epsilon_0 \left|\mathbf{E}\right|^2 = \mu_0 \left|\mathbf{H}\right|^2 \tag{2.287}$$

is the energy density.

Equation (2.287) is derived by considering equations (2.239, 2.240, and 2.285). For a plane wave with the electric field, $\mathbf{E}$, along the x direction, the Poynting vector, $\mathbf{S}$, is expressed as

$$\begin{aligned} \mathbf{S} &= \mathbf{E}\times\mathbf{H} = E_x H_x \hat{\boldsymbol{\kappa}} \\ &= \sqrt{\frac{\epsilon_0}{\mu_0}} \mid \mathbf{E} \mid^2 \hat{\boldsymbol{\kappa}} = \epsilon_0 c \mid \mathbf{E} \mid^2 \hat{\boldsymbol{\kappa}} \\ &= \sqrt{\frac{\mu_0}{\epsilon_0}} \mid \mathbf{H} \mid^2 \hat{\boldsymbol{\kappa}} = \mu_0 c \mid \mathbf{H} \mid^2 \hat{\boldsymbol{\kappa}} \end{aligned} \tag{2.288}$$

By combining equations (2.287 and 2.288), we find

$$\mathbf{S} = \frac{c}{\sqrt{\epsilon_0 \mu_0}} \frac{\boldsymbol{\kappa}}{\omega} w = \frac{\boldsymbol{\kappa}}{\omega} vw \tag{2.289}$$

The Poynting vector represents the flow of energy, with respect to both its magnitude and direction of propagation. Expressing $\mathbf{E}$ and $\mathbf{H}$ in complex terms, the time-averaged flux of energy is given by the real part of the Poynting vector

$$\mathbf{S} = \frac{1}{2}(\mathbf{E}\times\mathbf{H}^*) \tag{2.290}$$

in which * represents the complex conjugate.

Thus, we may write

$$\mathbf{S} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}} \frac{\boldsymbol{\kappa}}{\omega} (\mathbf{E}\cdot\mathbf{E}^*) \tag{2.291}$$

© 2015 by Taylor & Francis Group, LLC

In order to described the strength of a wave, the amount of energy carried by the wave in unit time across unit area perpendicular to the direction of propagation is used. This quantity, known as intensity (see Section 1.5), I, of the wave, is given in equation (2.286). The Poynting vector, $\mathbf{S}$, in terms of spherical coordinates is written as

$$\mathbf{S} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\frac{\boldsymbol{\kappa}}{\omega}\left(E_\theta E_\theta^* + E_\phi E_\phi^*\right) \tag{2.292}$$

The quantity within the parentheses represents the total intensity of the wave field, known as first Stokes parameter S_0. Thus, the Poynting vector is directly proportional to first Stokes parameter (see Section 8.4.2).

2.6.2 Harmonic Time Dependence

Maxwell's equations (see Section 2.4) for an electromagnetic field with general time dependence are simplified by specifying a field with harmonic dependence ([166], [160]). The harmonic time dependent electromagnetic fields are given by

$$\mathbf{E}(\mathbf{r},t) = \Re\left[\mathbf{E}_0(\mathbf{r},\omega)e^{i\omega t}\right] \tag{2.293}$$

$$\mathbf{B}(\mathbf{r},t) = \Re\left[\mathbf{B}_0(\mathbf{r},\omega)e^{i\omega t}\right] \tag{2.294}$$

in which $\mathbf{E}_0$ is a complex vector with Cartesian rectangular components

$$\begin{aligned}\mathbf{E}_{0x} &= a_1(\mathbf{r},\omega)e^{i\psi_1(\mathbf{r},\omega)}\\ \mathbf{E}_{0y} &= a_2(\mathbf{r},\omega)e^{i\psi_2(\mathbf{r},\omega)}\\ \mathbf{E}_{0z} &= a_3(\mathbf{r},\omega)e^{i\psi_3(\mathbf{r},\omega)}\end{aligned} \tag{2.295}$$

where $a_j(\mathbf{r},\omega)$ is the amplitude of the electric wave and $j = 1, 2, 3$.

For a homogeneous plane wave, the amplitudes, $a_j(\mathbf{r},\omega)$, are constant and the phase functions are $\psi_j(\mathbf{r},\omega) = \boldsymbol{\kappa}\cdot\mathbf{r} - \delta_j$, in which δ_j are the phase constants, which specify the state of polarization and $\boldsymbol{\kappa}$ the propagation vector. Each component of the vector phasor $\mathbf{E}_0$ has a modulus a_j and argument ψ_j, which depend on the position $\mathbf{r}$ and the parameter ω. The unit of this vector phasor $E_0(\mathbf{r},\omega)$ for harmonic time dependence is $\mathrm{V.m^{-1}}$. By differentiating equation (2.293) with respect to the temporal or spatial variables, Maxwell's equation (2.278) turns out to be

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{E}(\mathbf{r},t) &= \boldsymbol{\nabla}\times\Re\left[\mathbf{E}_0(\mathbf{r},\omega)e^{i\omega t}\right]\\ &= -\frac{\partial}{\partial t}\Re\left[\mathbf{E}_0(\mathbf{r},\omega)e^{i\omega t}\right]\\ &= \Re\left[-i\omega\mathbf{B}_0(\mathbf{r},\omega)e^{i\omega t}\right]\end{aligned} \tag{2.296}$$

By rearranging equation (2.296), we get

$$\boldsymbol{\nabla}\times\mathbf{E}_0(\mathbf{r},\omega)e^{i\omega t} = -i\omega\mathbf{B}_0(\mathbf{r})e^{i\omega t} \tag{2.297}$$

© 2015 by Taylor & Francis Group, LLC

or

$$\boldsymbol{\nabla}\times\mathbf{E}_0(\mathbf{r},\omega) = -i\omega\mathbf{B}_0(\mathbf{r},\omega) \tag{2.298}$$

Similarly, other Maxwell's equations may also be derived as

$$\begin{aligned}
\boldsymbol{\nabla}\times\mathbf{H}_0(\mathbf{r},\omega) &= \mathbf{J}_0(\mathbf{r},\omega) + i\omega\mathbf{D}_0(\mathbf{r},\omega) && (2.299)\\
\boldsymbol{\nabla}\cdot\mathbf{D}_0(\mathbf{r},\omega) &= \rho(\mathbf{r},\omega) && (2.300)\\
\boldsymbol{\nabla}\cdot\mathbf{B}_0(\mathbf{r},\omega) &= 0 && (2.301)\\
\boldsymbol{\nabla}\cdot\mathbf{J}_0(\mathbf{r},\omega) &= -i\omega\rho(\mathbf{r},\omega) && (2.302)
\end{aligned}$$

Equations (2.299−2.302) are known as Maxwell's equations for the frequency domain. Maxwell's equations for the complex vector phasors, $\mathbf{E}_0(\mathbf{r},\omega)$, $\mathbf{B}_0(\mathbf{r},\omega)$, etc., are applied to electromagnetic systems in which the constitutive relations for all materials are time-invariant and linear. Maxwell's equations with a cosinusoidal excitation are solved to obtain the vector phasors for the electromagnetic field $\mathbf{E}(\mathbf{r},t)$, $\mathbf{B}(\mathbf{r},t)$. For harmonic time dependence, $\mathbf{E}(\mathbf{r},t) = \Re[\mathbf{E}_0(\mathbf{r},\omega)e^{i\omega t}]$, the Hermitian magnitude of a complex vector is $|\mathbf{E}_0| = [\mathbf{E}_0\cdot\mathbf{E}_0^*]^{1/2}$. If the electromagnetic field is harmonic in time, the instantaneous rate at which energy is exchanged between the field and the mechanical motion of the charge is the product of

$$\mathbf{E}(\mathbf{r},t)\cdot\mathbf{J}(\mathbf{r},t) = \Re\left[\mathbf{E}_0(\mathbf{r},\omega)e^{i\omega t}\right]\cdot\Re\left[\mathbf{J}_0(\mathbf{r},\omega)e^{i\omega t}\right] \tag{2.303}$$

The real terms are written as

$$\begin{aligned}
\Re\left[\mathbf{E}_0(\mathbf{r},\omega)e^{i\omega t}\right] &= \frac{1}{2}\left[\mathbf{E}_0(\mathbf{r},\omega)e^{i\omega t} + \mathbf{E}_0^*(\mathbf{r},\omega)e^{-i\omega t}\right]\\
\Re\left[\mathbf{J}_0(\mathbf{r},\omega)e^{i\omega t}\right] &= \frac{1}{2}\left[\mathbf{J}_0(\mathbf{r},\omega)e^{i\omega t} + \mathbf{J}_0^*(\mathbf{r},\omega)e^{-i\omega t}\right] && (2.304)
\end{aligned}$$

The scalar product of these two terms provides

$$\begin{aligned}
\mathbf{E}(\mathbf{r},t)\cdot\mathbf{J}(\mathbf{r},t) &= \frac{1}{4}\left[\mathbf{E}_0(\mathbf{r},\omega)\cdot\mathbf{J}_0^*(\mathbf{r},\omega) + \mathbf{E}_0^*(\mathbf{r},\omega)\cdot\mathbf{J}_0(\mathbf{r},\omega)\right.\\
&\quad \left.+\mathbf{E}_0(\mathbf{r},\omega)\cdot\mathbf{J}_0(\mathbf{r},\omega)e^{2i\omega t} + \mathbf{E}_0^*(\mathbf{r},\omega)\cdot\mathbf{J}_0^*(\mathbf{r},\omega)e^{-2i\omega t}\right]\\
&= \frac{1}{2}\left\{\Re\left[\mathbf{E}_0(\mathbf{r},\omega)\cdot\mathbf{J}_0^*(\mathbf{r})\right] + \Re\left[\mathbf{E}_0(\mathbf{r},\omega)\cdot\mathbf{J}_0(\mathbf{r},\omega)e^{2i\omega t}\right]\right\}
\end{aligned} \tag{2.305}$$

Since the optical frequencies are very large, one can observe their time average over a period of oscillation, $T = 2\pi/\omega$. The time average over a time that is large compared with the inverse frequency of the product of the two harmonic time-independent functions $\mathbf{a}$ and $\mathbf{b}$ of the same frequency is given by

$$\begin{aligned}
\langle\mathbf{a}(t)\cdot\mathbf{b}(t)\rangle &= \frac{1}{T}\int_0^T \frac{1}{4}\left[\mathbf{a}e^{i\omega t} + \mathbf{a}^*e^{-i\omega t}\right]\cdot\left[\mathbf{b}e^{i\omega t} + \mathbf{b}^*e^{-i\omega t}\right]dt\\
&= \frac{1}{2}\Re\left(\mathbf{a}\cdot\mathbf{b}^*\right) = \frac{1}{2}\Re\left(\mathbf{a}^*\cdot\mathbf{b}\right) && (2.306)
\end{aligned}$$

© 2015 by Taylor & Francis Group, LLC

where $\langle\ \rangle$ stands for the time average.

Hence, the time average of the product $\langle \mathbf{E}(\mathbf{r},t)\cdot\mathbf{J}(\mathbf{r},t)\rangle$ is expressed as

$$\langle \mathbf{E}(\mathbf{r},t)\cdot\mathbf{J}(\mathbf{r},t)\rangle = \frac{1}{2}\Re\left(\mathbf{E}_0(\mathbf{r},\omega)\cdot\mathbf{J}_0^*(\mathbf{r},\omega)\right) \tag{2.307}$$

Similarly, the time average value of the Poynting vector product may also be derived as

$$\begin{aligned}
\langle \mathbf{S}(\mathbf{r},t)\rangle &= \frac{1}{T}\int_0^T \left[\mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\right] dt \\
&= \frac{1}{T}\int_0^T \frac{1}{4}\left[\mathbf{E}_0(\mathbf{r},\omega)\times\mathbf{H}_0^*(\mathbf{r},\omega) + \mathbf{E}_0^*(\mathbf{r},\omega)\times\mathbf{H}_0(\mathbf{r},\omega)\right. \\
&\quad \left. +\mathbf{E}_0(\mathbf{r},\omega)\times\mathbf{H}_0(\mathbf{r},\omega)e^{2i\omega t} + \mathbf{E}_0^*(\mathbf{r},\omega)\times\mathbf{H}_0^*(\mathbf{r},\omega)e^{-2i\omega t}\right] dt \\
&\simeq \frac{1}{4}\left[\mathbf{E}_0(\mathbf{r},\omega)\times\mathbf{H}_0^*(\mathbf{r},\omega) + \mathbf{E}_0^*(\mathbf{r},\omega)\times\mathbf{H}_0(\mathbf{r},\omega)\right] \\
&= \frac{1}{2}\Re\left[\mathbf{E}_0(\mathbf{r},\omega)\times\mathbf{H}_0^*(\mathbf{r},\omega)\right] = \Re[\mathbf{S}_c(\mathbf{r},\omega)]
\end{aligned} \tag{2.308}$$

Thus, the complex Poynting vector is deduced as

$$\mathbf{S}_c(\mathbf{r},\omega) = \frac{1}{2}[\mathbf{E}_0(\mathbf{r},\omega)\times\mathbf{H}_0^*(\mathbf{r},\omega)] \tag{2.309}$$

The real part of this Poynting vector is known as the time average of the Poynting vector. The law of conservation of energy takes a simple form. The complex Poynting theorem is given by

$$\int_{\mathtt{V}}\left(\frac{1}{2}\mathbf{E}_0\cdot\mathbf{J}_0^*\right) d\mathtt{V} - i\omega\int_{\mathtt{V}}\frac{1}{2}\left(\mathbf{E}_0\cdot\mathbf{D}_0^* - \mathbf{H}_0^*\cdot\mathbf{B}_0\right) d\mathtt{V} + \oint_{\mathtt{S}}\mathbf{S}_c\cdot d\mathbf{S} = 0 \tag{2.310}$$

For nonconducting medium ($\sigma = 0$), where no mechanical work is done, the time average of equation (2.252) turns out to be

$$\boldsymbol{\nabla}\cdot\langle\mathbf{S}(\mathbf{r},t)\rangle = 0 \tag{2.311}$$

By integrating equation (2.311) over an arbitrary volume which contains no absorber or radiator of energy, we obtain, after applying Gauss' theorem,

$$\oint_{\mathtt{S}}\langle\mathbf{S}(\mathbf{r},t)\rangle\cdot\hat{\mathbf{n}}d\mathbf{S} = 0 \tag{2.312}$$

in which $\hat{\mathbf{n}}$ is the outward normal to the surface.

Thus, the averaged total flux of energy through any closed surface is zero. The time average of the electric energy density is derived as

$$\begin{aligned}
\langle w_e\rangle &= \frac{1}{T}\int_0^T \frac{\epsilon_0}{2}\mathbf{E}^2 dt \\
&= \frac{\epsilon_0}{2}\frac{1}{T}\int_0^T \frac{1}{4}\left[\mathbf{E}_0^2 e^{2i\omega t} + \mathbf{E}_0\cdot\mathbf{E}_0^* + {\mathbf{E}_0}^{*2} e^{-2i\omega t}\right] dt
\end{aligned} \tag{2.313}$$

© 2015 by Taylor & Francis Group, LLC

Since T is assumed to be large, the integrals involving the exponentials are neglected. Therefore, we get

$$\langle w_e \rangle = \frac{\epsilon_0}{4}\mathbf{E}_0 \cdot \mathbf{E}_0^* = \frac{\epsilon_0}{4}|\mathbf{E}_0|^2 \tag{2.314}$$

Similarly, the time average of the magnetic energy density is also derived as

$$\langle w_m \rangle = \frac{\mu_0}{4}\mathbf{H}_0 \cdot \mathbf{H}_0^* = \frac{\mu_0}{4}|\mathbf{H}_0|^2 \tag{2.315}$$

Thus, the difference in the time-average energies stored in the electric and magnetic fields with the volume

$$\langle w_e \rangle - \langle w_m \rangle = \frac{\epsilon_0}{4}|\mathbf{E}_0|^2 - \frac{\mu_0}{4}|\mathbf{H}_0|^2 = 0 \tag{2.316}$$

Hence $\langle w_e \rangle = \langle w_m \rangle$. The total energy W is given by

$$W = 2\langle w_e \rangle = 2\langle w_m \rangle$$

2.6.3 Fourier Transform of Harmonic Equations

The Fourier transform (see Section 1.3.3) of an electric field is expressed as

$$\widehat{\mathbf{E}}_0(\mathbf{r}, \omega) = \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r}, t) e^{-i\omega t} dt \tag{2.317}$$

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{\mathbf{E}}_0(\mathbf{r}, \omega) e^{i\omega t} d\omega \tag{2.318}$$

The Fourier transform of $\widehat{\mathbf{E}}_0(\mathbf{r}, \omega)$ is a complex function of the variable ω and has the units of the electric field $\mathrm{V.m^{-1}}$ per unit frequency, i.e., $\mathrm{V.m^{-1}.Hz^{-1}}$. By invoking the principle of Fourier transform of a temporal derivative of a function, $df(t)/dt \leftrightarrow i\omega\widehat{f}(\omega)$, the curl of equation (2.318) is applied on Maxwell's equation; thus,

$$\begin{aligned} \boldsymbol{\nabla} \times \widehat{\mathbf{E}}_0(\mathbf{r}, \omega) &= \int_{-\infty}^{\infty} \boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r}, t) e^{-i\omega t} dt \\ &= \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial t}\mathbf{B}(\mathbf{r}, t)\right] e^{-i\omega t} dt \end{aligned} \tag{2.319}$$

The Fourier transform of the magnetic field is given by

$$\widehat{\mathbf{B}}_0(\mathbf{r}, \omega) = \int_{-\infty}^{\infty} \mathbf{B}(\mathbf{r}, t) e^{-i\omega t} dt \tag{2.320}$$

therefore, equation (2.319) turns out to be

$$\boldsymbol{\nabla} \times \widehat{\mathbf{E}}_0(\mathbf{r}, \omega) = -i\omega\widehat{\mathbf{B}}_0(\mathbf{r}, \omega) \tag{2.321}$$

© 2015 by Taylor & Francis Group, LLC

Similar operations may be applied with other Maxwell's equations in order to derive their Fourier transforms as well

$$\boldsymbol{\nabla}\times\widehat{\mathbf{H}}_0(\mathbf{r},\omega) = \widehat{\mathbf{J}}_0(\mathbf{r},\omega) + i\omega\widehat{\mathbf{D}}_0(\mathbf{r},\omega) \quad (2.322)$$

$$\boldsymbol{\nabla}\cdot\widehat{\mathbf{D}}_0(\mathbf{r},\omega) = \widehat{\rho}(\mathbf{r},\omega) \quad (2.323)$$

$$\boldsymbol{\nabla}\cdot\widehat{\mathbf{B}}_0(\mathbf{r},\omega) = 0 \quad (2.324)$$

$$\boldsymbol{\nabla}\cdot\widehat{\mathbf{J}}_0(\mathbf{r},\omega) = -i\omega\widehat{\rho}(\mathbf{r},\omega) \quad (2.325)$$

These are Maxwell's equations for the Fourier transform of the electromagnetic fields, $\mathbf{E}_0(\mathbf{r},\omega)$, $\mathbf{B}_0(\mathbf{r},\omega)$, etc.

Let us examine the integral of the complex Poynting vector

$$\frac{2}{\pi}\int_0^\infty \hat{\mathbf{n}}\cdot\Re[\mathbf{S}_c(\mathbf{r},\omega)]d\omega = \frac{2}{\pi}\int_0^\infty \hat{\mathbf{n}}\cdot\left\{\frac{1}{2}\Re\left[\widehat{\mathbf{E}}_0(\mathbf{r},\omega)\times\widehat{\mathbf{H}}_0^*(\mathbf{r},\omega)\right]\right\}d\omega \quad (2.326)$$

Let us introduce the Parseval theorem (or Power theorem; see Section 8.2.2), so that

$$\begin{aligned}\int_{-\infty}^\infty \hat{\mathbf{n}}\cdot\mathbf{S}(\mathbf{r},t)dt &= \int_{-\infty}^\infty \hat{\mathbf{n}}\cdot[\mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)]dt \\ &= \frac{2}{\pi}\int_0^\infty \hat{\mathbf{n}}\cdot\left\{\frac{1}{2}\Re\left[\widehat{\mathbf{E}}_0(\mathbf{r},\omega)\times\widehat{\mathbf{H}}_0^*(\mathbf{r},\omega)\right]\right\}d\omega \\ &= \frac{2}{\pi}\int_0^\infty \hat{\mathbf{n}}\cdot\left\{\Re\left[\widehat{\mathbf{S}}_c(\mathbf{r},\omega)\right]\right\}d\omega \quad (2.327)\end{aligned}$$

The LHS of equation (2.327) is the total electromagnetic energy passing through a unit area of surface with the unit normal $\hat{\mathbf{n}}$, while the integrand on the RHS

$$\frac{2}{\pi}\hat{\mathbf{n}}\cdot\left\{\Re\left[\widehat{\mathbf{S}}_c(\mathbf{r},\omega)\right]\right\}$$

is the energy passing through a unit area of this surface per unit frequency, $\mathrm{J.m^{-2}.Hz^{-1}}$, and is known as the energy spectral density[44].

2.7 Antennas

Thus far, we have dealt with the propagation of the electromagnetic wave in free space where we assumed the solution of the source-free wave equation. However, the sources of waves are extremely important. In order to transmit power, we generate the waves at the proper level and frequencies, and couple the energy into the appropriate domain. This coupling is done by what is called an antenna([109], [136], [171]). It is basically a transducer which provides a transition from a guided wave on a transmission line to a free space wave

© 2015 by Taylor & Francis Group, LLC

or receives electromagnetic waves. A directional antenna system pushes the lobes in a certain direction causing the energy to be moved and condensed in a particular area (very little energy is in its rear side). Antennas have two complementary functions:

- converting electromagnetic waves into voltage and current used by a circuit and
- converting voltage and current into electromagnetic waves which are transmitted into space.

Signals are transmitted through space by electromagnetic waves consisting of electric fields measured in $V.m^{-1}$ and magnetic fields measured in $A.m^{-1}$.

2.7.1 Significance of Antenna Shape

Depending on the type of field being detected, the antenna takes on a particular construction. Antennas designed to pick up electric fields, such as the antenna of Figure 2.19a, are made with rods and plates while antennas made to pick up magnetic fields, as in Figure 2.19b, are made from loops of wire. Sometimes parts of electric circuits may have characteristics that unintentionally make them antennas. The electromagnetic circuit is concerned with reducing the probability of these unintentional antennas injecting signals into their circuits or influencing other circuits.

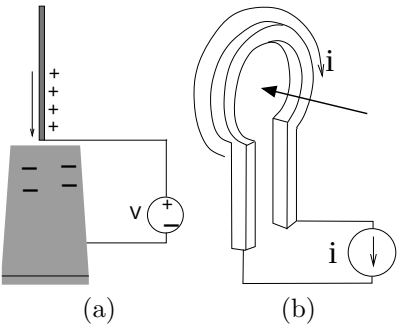


FIGURE 2.19: Antennas for picking up (a) electric and (b) magnetic fields.

In the antennas that are made of loops of wire, the current in a coil produces a magnetic field through the coil. Similarly, a magnetic field, passing through the coil, produces an electric current in the wire of the coil. The ends of the loop antenna are attached to a receiving circuit through which this induced current flows as the loop antenna detects the magnetic field. The magnetic fields are, in general, directed perpendicular to the direction of their propagation so the plane of the loop should be aligned parallel to the direction of the wave propagation to detect the field. Some types of electric field an-

© 2015 by Taylor & Francis Group, LLC

tennas are bi-conical, horn, and micro-strip. Generally, antennas that radiate electric fields have two components insulated from each other.

2.7.2 Radiation of Electromagnetic Wave by Antenna

The simplest electric field antenna is the dipole antenna whose very name implies its two-component nature. The two-conductor elements act like the plates of a capacitor (see Section 2.1.8) with the field between them projecting out into space rather than being confined between the plates. On the other hand, magnetic field antennas act as inductors (see Section 2.3.2). The inductor fields are projected out into space rather than being confined to a closed magnetic circuit. The categorization of antennas in this fashion is somewhat artificial since the actual mechanism of radiation involves both electric and magnetic fields.

In a simple parallel plate capacitor as shown in Figure 2.20a, the electric field that occurs when a charge is placed on each of the plates is contained between the plates. If the plates are spread apart so that they lie in the same plane, the electric field between the plates extends out into space. The same process occurs with an electric field dipole antenna as described in Figure 2.21a. Charges on each part of the antenna produce a field into space between the two halves of the antenna. There is an intrinsic capacitance between the two rods of the dipole antenna as depicted in Figure 2.20b. Current is required to charge the dipole rods. The current in each part of the antenna flows in the same direction. This condition is important since it results in radiation.

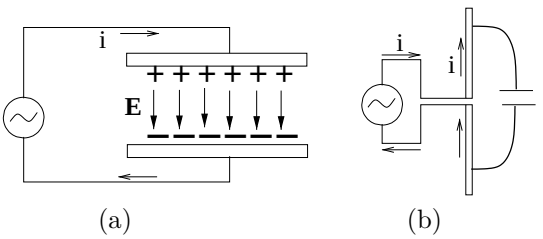


FIGURE 2.20: (a) Capacitor circuit and (b) dipole showing intrinsic capacitance and charging current.

The antenna radiation pattern is a measure of its power or radiation distribution with respect to a particular type of coordinate. This power variation as a function of the arrival angle is observed in the far-field (see Section 1.7.6) of the antenna. In general, we consider spherical coordinates as the ideal antenna is supposed to radiate in a spherically symmetrical pattern. However, antennas in practice are not omnidirectional but have a radiation maximum along one particular direction. For instance, a dipole antenna is a broadside antenna wherein the maximum radiation occurs along the axis of the antenna. In the case of the doughnut-shaped (or toroidal) radiation pattern, along the

© 2015 by Taylor & Francis Group, LLC

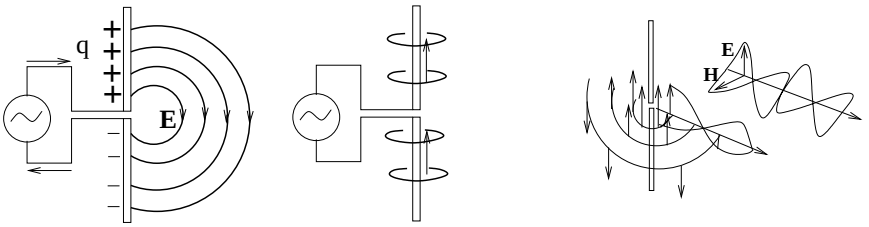


FIGURE 2.21: (a) Electric field, (b) magnetic field, and (c) transverse electromagnetic wave from dipole.

z axis corresponding to the radiation directly overhead the antenna, very little power is transmitted. In the x, y plane (perpendicular to the z axis), the radiation is maximum. The beam solid angle of a doughnut function, with $P(\theta, \phi) = \sin^2 \theta$ (the pattern of a dipole) is

$$\Omega_a = \int P(\theta, \phi) d\Omega = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \sin^2 \theta \sin \theta d\theta d\phi = 2\pi \int_0^{\pi} \sin^3 d\theta = \frac{8\pi}{3}$$

with

$$\int_0^{\pi} \sin^3 \theta d\theta = \int_0^{\pi} \left(\frac{3}{4} \sin \theta - \frac{1}{4} \sin 3\theta \right) d\theta = \frac{4}{3}.$$

As the signal applied to the two halves of the antenna oscillates, the field keeps reversing and sends out waves into space.

The principle of an antenna is that it must provide a time-dependent current, which, in turn, generates the electric and magnetic fields (see Figure 2.21a and 2.21b). When time-dependent electric and magnetic fields exist, power is generated and propagated. Equation (2.14; see Section 2.1.3.2) is the electrostatic definition of dipole, and because the charges were assumed to be constant, independent of time, this structure produces an electric field not a magnetic field. For a dipole to produce a magnetic field, it must produce a time-dependent current. Thus, for a dipole to serve as an antenna we require two conditions: (i) a current must flow in the dipole and (ii) the current must be time independent. These conditions to be satisfied for a structure to act as an antenna can be fulfilled by the following cases shown in Figure 2.22. Figure 2.22a displays the electrostatic dipole where two point charges are placed a distance, Δl, apart; one charge is negative and the other is positive. These charges are time dependent. We shall assume a sinusoidal time dependency (see Section 2.4.5), but any other time dependency may be used. Let us consider a simple dipole antenna, called the Hertzian dipole or point dipole antenna (see Figure 2.22b). In this, two point charges are connected through a thin conducting wire; it is infinitesimally short with uniform current distribution. A time-dependent current can now flow back and forth between

© 2015 by Taylor & Francis Group, LLC

these charges. It is important to mention that the total charge at any time must be zero, as required by the law of conservation of charge.

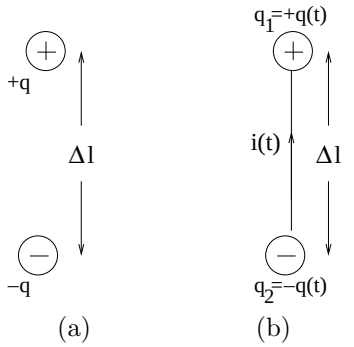


FIGURE 2.22: (a) Electrostatic dipole and (b) Hertzian dipole.

Since the radiation from a dipole depends on frequency, f, we consider a driving current, I, varying sinusoidally with angular frequency $\omega(= 2\pi f)$. Let us generate time-dependent charges as

$$\begin{aligned} q_1 &= q_0 \sin \omega t \rightarrow q_0 e^{i\omega t} \\ q_2 &= -q_0 \sin \omega t \rightarrow q_0 e^{i\omega t} \end{aligned} \tag{2.328}$$

With these charges, the current in the wire between charges is

$$I = I_0 e^{i\omega t} = \frac{\partial q_1}{\partial t} = -\frac{\partial q_2}{\partial t} \tag{2.329}$$

where I_0 is the peak current going into each half of the dipole.

In terms of charges, we may write the current as

$$I_0 = i\omega q_1 = -i\omega q_2 = i\omega q_0 \tag{2.330}$$

The dipole moment $\mathbf{p}$, which is a phasor, may now be deduced. With the dipole at the origin and current in the z direction as shown in Figure 2.22b, the electric or Hertzian dipole moment is

$$\mathbf{p} = \mathbf{z} q_0 \Delta l e^{i\omega t} \tag{2.331}$$

in which l is the total length of the dipole.

For time-varying current, let us replace i by $i(t)$. The solution seems to be simple, but requires deriving the Biot−Savart law (see Section 2.2.3), vector potential (see Section 2.2.4.2), as well as the displacement current (see Section 2.4.2.2). This solution is applicable to static and slowly varying fields but not to propagating waves. The time-varying current generates the magnetic vector potential everywhere in space instantaneously. Since the solution must

© 2015 by Taylor & Francis Group, LLC

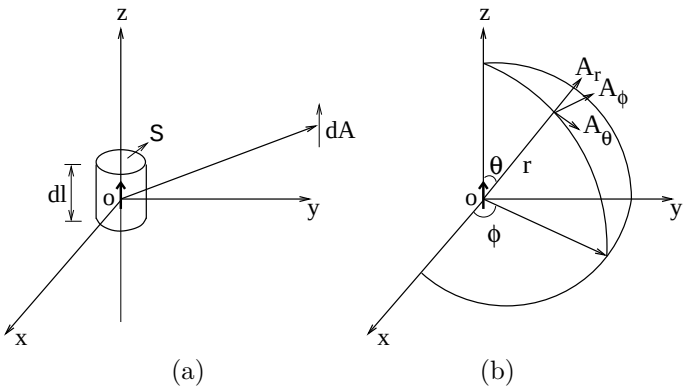


FIGURE 2.23: (a) Retarded vector magnetic potential and (b) resolved part of vector potential.

also obey the wave equation, the wave generated at the source propagates outward from the source in all directions with equal speed and a velocity $v = 1/\sqrt{\mu\epsilon}$ (see Section 2.6). If an observer is at a distance r from the source, a wave generated at the source will arrive at the observer after a time $t = r/v$. Thus, the wave that arrives at a distance r is a retarded wave: the vector potential becomes a function of t as

$$\mathbf{A}(t) = \frac{1}{4\pi} \int \frac{i(t - r/v)dl}{r} \tag{2.332}$$

and the retarded scalar potential V is

$$V(R,t) = \frac{1}{4\pi\epsilon} \int_{\mathbb{V}'} \frac{\rho(t - R/v)dl}{R} \tag{2.333}$$

With the definition of the retarded magnetic potential (see Figure 2.23a), we extend the analysis to derive the magnetic potential (see Section 2.2.4) due to a time-varying current, $i_0 \cos\omega t$. The retarded magnetic vector potential for a conductor of infinitesimal cross-section $d\mathbb{S}$ and length dl is given by

$$\begin{aligned} d\mathbf{A} &= \frac{[i_0 \cos\omega(t - r/v)]\, d\mathbf{l}}{4\pi r} \\ &= \frac{[i_0(\cos\omega t')]\, d\mathbf{l}}{4\pi r} \end{aligned} \tag{2.334}$$

The vector magnetic potential due to the finite size conductor is obtained, integrating equation (2.334) over the whole cross-section of the conductor:

$$\mathbf{A} = \frac{\cos\omega t' d\mathbf{l}}{4\pi r} \int \frac{i_0}{r} d\mathbb{S} \tag{2.335}$$

© 2015 by Taylor & Francis Group, LLC

To evaluate this integral, if $r \gg$ diameter of current carrying conductor, r may be regarded as being constant; and thus,

$$\mathbf{A} = \frac{[I_0 \cos \omega(t - r/v)]\, d\mathbf{l}}{4\pi r} \tag{2.336}$$

where $I_0 \cos \omega t$ denotes the total current in the finite-sized conductor and I_0 is obtained by integrating i over the cross-section S of the conductor.

This $\mathbf{A}$, which is in the same direction as $d\mathbf{l}$ can be resolved into the three components in spherical co-ordinates as

$$A_r = A \cos\theta \qquad A_\theta = -A \sin\theta \qquad A_\phi = 0$$

2.7.2.1 Electric and Magnetic Fields of Oscillating Hertzian Dipole

The short element that is described in the preceding section carrying time-varying current can be thought of as consisting of equal and opposite charges along the z axis situated symmetrically with respect to the origin, O (see Figure 2.23b). The magnitude of the charge at every point along dl varies sinusoidally with time. Such a system is, therefore, described as an oscillating Hertzian dipole[134]. In order to find the electric and magnetic field intensities in the region surrounding the dipole antenna, we make use of the retarded vector. Let us write

$$\mathbf{H} = \mathbf{H}_r \hat{\mathbf{n}}_r + \mathbf{H}_\theta \hat{\mathbf{n}}_\theta + \mathbf{H}_\phi \hat{\mathbf{n}}_\phi = \boldsymbol{\nabla} \times \mathbf{A} \tag{2.337}$$

Using the curl equation in spherical coordinates (see equation B.31; Appendix B.2.4), we find

$$\mathbf{H}_r = \frac{1}{r \sin\theta} \left[\frac{\partial}{\partial\theta}(\sin\theta A_\phi) - \frac{\partial A_\theta}{\partial\phi} \right] = 0 \tag{2.338}$$

$$\mathbf{H}_\theta = \frac{1}{r} \left[\frac{1}{\sin\theta} \frac{\partial A_r}{\partial\phi} - \frac{\partial}{\partial r}(r A_\phi) \right] = 0 \tag{2.339}$$

$$\begin{aligned} \mathbf{H}_\phi &= \frac{1}{r} \left[\frac{\partial}{\partial r}(r A_\theta) - \frac{\partial A_r}{\partial\theta} \right] \\ &= \frac{I_0 dl}{4\pi} \left[\frac{1}{r} \frac{\partial}{\partial r} \left\{ -\sin\theta \cos\omega \left(t - \frac{r}{v} \right) \right\} \right. \\ &\quad \left. - \frac{1}{r} \frac{\partial}{\partial\theta} \left\{ \frac{\cos\theta}{r} \cos\omega \left(t - \frac{r}{v} \right) \right\} \right] \end{aligned} \tag{2.340}$$

It is interesting to witness that $\mathbf{H}_r$- and $\mathbf{H}_\theta$-components (equations 2.338 and 2.339) are zero, but the $\mathbf{H}_\phi$-component (equation 2.340) consists of two terms:

- the first term varying inversely with r falls off less rapidly and is dominant in the so-called the radiation or distant field and

© 2015 by Taylor & Francis Group, LLC

- the second varies inversely with r^2, which is the Biot−Savart type of expression (see Section 2.2.3).

The second term decreases more rapidly with distance, and at large distances, it becomes negligibly small, so that for large values of r, the term remains. The expressions for the $\mathbf{E}_r$-, $\mathbf{E}_\theta$-, and $\mathbf{E}_\phi$-components of the electric field by employing the components of the curl equation (see Appendix B.2.4) are respectively given by

$$\begin{aligned}
\mathbf{E}_r &= \frac{1}{\epsilon}\int (\boldsymbol{\nabla}\times\mathbf{H})_r dt \\
&= \frac{1}{\epsilon}\int \left[\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\mathbf{H}_\phi) - \frac{\partial\mathbf{H}_\theta}{\partial\phi}\right] dt \\
&= \frac{I_0 dl}{4\pi\epsilon}\int \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left\{\sin^2\theta\left(\frac{-\omega}{vr}\sin\omega t'\right) + \sin^2\theta\frac{\cos\omega t'}{r^2}\right\} dt \\
&= \frac{2\cos\theta I_0 dl}{4\pi\epsilon}\left[\frac{\cos\omega t'}{vr^2} + \frac{\sin\omega t'}{\omega r^3}\right] \qquad (2.341) \\
\mathbf{E}_\theta &= \frac{1}{\epsilon}\int (\boldsymbol{\nabla}\times\mathbf{H})_\theta dt \\
&= \frac{1}{\epsilon}\int \left[\frac{1}{r\sin\theta}\frac{\partial\mathbf{H}_r}{\partial\phi} - \frac{1}{r}\frac{\partial}{\partial r}(r\mathbf{H}_\phi)\right] dt \\
&= \frac{\sin\theta I_0 dl}{4\pi\epsilon}\left[\frac{-\omega\sin\omega t'}{rv^2} + \frac{\cos\omega t'}{r^2 v} + \frac{\sin\omega t'}{\omega r^3}\right] \qquad (2.342) \\
\mathbf{E}_\phi &= \frac{1}{\epsilon}\int (\boldsymbol{\nabla}\times\mathbf{H})_\phi dt \\
&= \frac{1}{\epsilon}\int \left[\frac{1}{r}\frac{\partial}{\partial r}(r\mathbf{H}_\theta) - \frac{\partial\mathbf{H}_r}{\partial\theta}\right] dt = 0 \qquad (2.343)
\end{aligned}$$

Equation (2.342) shows the radiation field term varying as $1/r$ and the induction field term (which is the electrodynamic version of the Biot−Savart law) varying as $1/r^2$. However, equation (2.341) does not have the term varying as $1/r$, which means that the $\mathbf{E}_r$-component of the field does not contribute to the radiation field; the induction field term varying as $1/r^2$ is present in $\mathbf{E}_r$. Moreover, both $\mathbf{E}_r$ and $\mathbf{E}_\theta$ have an additional term varying as $1/r^3$, which is a field component of an electrostatic dipole and is referred to as the electrostatic field term. Owing to the inverse cube factor, these terms decrease very fast with distance and are insignificant everywhere except very close to the origin. The induction, as well as radiation terms, would have equal amplitude at a critical distance, r_c, if

$$\frac{\omega}{vr_c} = \frac{1}{r_c^2} \qquad (2.344)$$

with $r_c = \frac{v}{\omega} = \frac{\lambda}{2\pi} \approx \frac{\lambda}{6}$, which implies that for distance greater than $\lambda/6$, the amplitude of radiation term dominates.

© 2015 by Taylor & Francis Group, LLC

2.7.2.2 Radiation Fields of a Hertzian Dipole Antenna

Since the electromagnetic fields at large distance from the antenna are important, let us confine the discussion to the radiation field. Using the relations, $v = 1/\sqrt{\mu\epsilon}$, $\omega/v = 2\pi/\lambda$, and $\eta = 120\pi$, a radiated wave moves through the atmosphere for which the characteristic impedance may be considered the same as that of free space. Thus, we get

$$\mathbf{E}_\theta = \frac{60\pi I_0 dl \sin\theta}{\lambda r}\left[-\sin\theta\left(t - \frac{r}{v}\right)\right] \tag{2.345}$$

$$\mathbf{H}_\phi = \frac{\mathbf{E}_\theta}{\eta} \tag{2.346}$$

while $\mathbf{E}_r$, $\mathbf{E}_\phi$, $\mathbf{H}_r$, and $\mathbf{H}_\theta$ are zero.

It is pertinent to note that the radiation wave propagates in an r direction and the $\mathbf{E}$ and $\mathbf{H}$ fields are perpendicular to the direction of propagation; the $\mathbf{E}$ and $\mathbf{H}$ fields are mutually perpendicular to each other as well (see Section 2.6.1). Thus, the radiated waves are transverse electromagnetic waves (TEM; see Figure 2.21c).

2.7.2.3 Radiation Resistance of an Antenna

The power, P, flowing through a circuit is defined as $P = Vi$, in which V is the voltage (energy per unit charge; see Section 2.1.7) and i the current (charge flow per unit time). It has dimensions of energy per unit time. As discussed in Section 2.1.9, the resistance $R = V/i$; hence, the power turns out to be $P = V^2/R$.

The radiated energy of an antenna can be derived by taking the time-averaged Poynting vector $\langle\mathbf{S}\rangle$ power per unit area (see Section 2.5.1) over one period. Let us assume that in spherical coordinates, $\mathbf{S}_r$, $\mathbf{S}_\theta$, and $\mathbf{S}_\phi$ are the components of the Poynting vector. In the case of the radiation term, the power flow is given by

$$\begin{aligned}\mathbf{E}_r &= \left(\frac{I_0 \sin\theta dl}{4\pi}\right)^2 \frac{\omega^2}{\epsilon v^3 r^2}\sin^2\omega t' \\ &= \left(\frac{I_0 \sin\theta dl}{4\pi}\right)^2 \frac{\omega^2}{\epsilon v^3 r^2}\left[\frac{1-\cos 2\omega t'}{2}\right]\end{aligned} \tag{2.347}$$

The part with $\cos 2\omega t'$ contributes zero energy flow over a complete cycle; hence, the constant term within the square bracket provides a constant flow of energy in the r direction as

$$\begin{aligned}\langle\mathbf{S}_r\rangle &= \left(\frac{I_0 \sin\theta dl}{4\pi}\right)^2 \frac{\omega^2}{\epsilon v^3 r^2} \times \frac{1}{2} \\ &= \frac{\eta}{2}\left(\frac{\omega I_0 \sin\theta dl}{4\pi v r}\right)^2 \quad \text{W.m}^{-2}\end{aligned} \tag{2.348}$$

© 2015 by Taylor & Francis Group, LLC

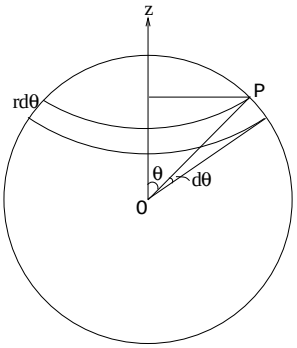


FIGURE 2.24: Illustration of obtaining the total radiated power.

in which $\epsilon v = 1/\eta$ and $\langle \cos^2(\omega t) \rangle = 1/2$.

As stated in Section 2.7.2, the power pattern is the angular distribution of radiated power (see Figure 2.24), which is often normalized to unity at the peak. The radiation from a short dipole has the same polarization (see Section 8.4) and the same doughnut-shaped power pattern as Larmor radiation from an accelerated charge because all of the charges in the dipole are being accelerated along one line much shorter than one wavelength. At the observer's site, the received power depends on the projected (perpendicular to the line of sight) length ($l \sin\theta$) of the dipole, in which l is the total length of the dipole. Thus, the electric field received is proportional to the apparent length of the dipole. The time-averaged total power radiated by an antenna in all directions is obtained by integrating the Poynting flux, $\langle \mathbf{S}_r \rangle$, over the surface of a sphere of radius $r \gg l$; thus, the time-averaged power radiated by the antenna is

$$\begin{aligned} \langle P \rangle &= \int \langle \mathbf{S}_r \rangle \cdot d\mathbf{S} \\ &= \frac{\eta_0}{2} \left[\frac{\omega I_0 dl}{4\pi v r} \right]^2 \int_0^\pi \sin^2\theta (2\pi r \sin\theta r d\theta) \\ &= \frac{\eta_0}{2} \left[\frac{\omega I_0 dl}{4v} \right]^2 \int_0^\pi \sin^3\theta d\theta \\ &= \frac{\eta_0 \omega^2 I_0^2 dl^2}{12\pi v^2} \end{aligned} \tag{2.349}$$

Using the relations $I_0 = \sqrt{2} I_{rms}$ and $\eta = 120\pi$, this equation (2.349) turns out to be

$$\langle P \rangle = 80\pi^2 \left(\frac{dl^2}{\lambda} \right) I_{rms}^2 \tag{2.350}$$

Since in an electrical circuit the power dissipated in a resistor R is $I_{rms}^2 R$

© 2015 by Taylor & Francis Group, LLC

(Ohm law; see Section 2.1.9), the radiation resistance is

$$\langle R \rangle = 80\pi^2 \left(\frac{dl}{\lambda}\right)^2 \tag{2.351}$$

2.7.2.4 Quarter-Wave Monopole and Half-Wave Dipole

Let us extend the Hertzian dipole of length dl to a practical dipole antenna of finite length L meters which is measured from one end of the antenna to the other end (see Figure 2.25a). In such a case the power is fed by a two-wire transmission line to the center of the antenna. In fact, we may think of the antenna as a continuation of a two-wire line, in which the line opens out, or its wires are bent at right angles to form the antenna. If we assume that the antenna is much less than λ, a simplified analysis can be made.

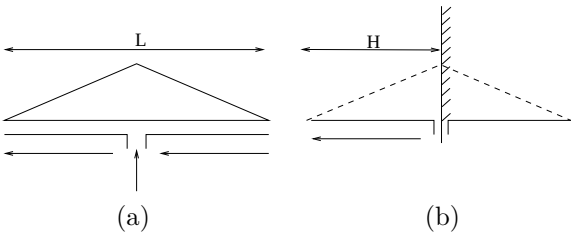


FIGURE 2.25: (a) Dipole and (b) monopole antennas.

In general, the dipoles are half-wave dipoles since they are resonant, which means that they provide a nearly resistive load to the transmitter. When each half of the dipole is $\lambda/4$ long, the standing-wave current is highest at the center and naturally falls to zero at the ends of the antenna. The power radiated from different parts of the antenna, therefore, varies. Since the antenna is much smaller than λ, we may assume that the current varies linearly from maximum value at the center to zero value at the ends. In such a situation, the average value is $I_0/2$, and it may be considered that the rms value of the current is $I_{rms}/2$. Thus,

$$\begin{aligned} \langle P \rangle_{dipole-L} &= 80\pi^2 \left(\frac{L}{\lambda}\right)^2 \left(\frac{I_{rms}}{2}\right)^2 \\ &= 20\pi^2 \left(\frac{L}{\lambda}\right)^2 I_{rms}^2 \end{aligned} \tag{2.352}$$

so that $\langle R \rangle_{dipole-L} = 20\pi^2 \left(\frac{L}{\lambda}\right)^2 = 200 \left(\frac{L}{\lambda}\right)^2 \Omega$.

We assume that at the frequency at which power is fed, we may take the method of images and use only the upper conductor of the transmission line and the upper half of the dipole antenna (see Figure 2.25b), the lower

© 2015 by Taylor & Francis Group, LLC

conductor being grounded. Hence, the power would be

$$\begin{aligned}\langle P\rangle_{monopole-H} &= \frac{1}{2}\left[20\pi^2\left(\frac{2H}{\lambda}\right)^2\right] I_{rms}^2 \\ &= 40\pi^2\left(\frac{2H}{\lambda}\right)^2 I_{rms}^2 \end{aligned} \tag{2.353}$$

so that $\langle R\rangle_{monopole-H} = 40\pi^2\left(\frac{2H}{\lambda}\right)^2 = 400\left(\frac{H}{\lambda}\right)^2\ \Omega$.

These derivations that we have used are justified by the fact that these equations are found to be valid in actual antennas for L up to $\lambda/4$ and H up to $\lambda/8$.

2.7.2.5 Power Gain of an Antenna

The gain in any direction is the ratio of the power density radiated in direction divided to the power density radiated by a perfect (loss-less) isotropic radiator having the same total accepted input power. The radiated power is dependent on both r and θ coordinates of P. For comparing the characteristics of different antennas, one is interested to know that the fraction of total power radiated by the antenna goes in a certain direction θ. Since the antenna does not radiate uniformly in all directions, we assume an imaginary isotropic antenna, which radiates the same total power $\langle P\rangle$ as the given antenna, but radiates it uniformly in all directions. The ratio of the radiation intensity in a particular direction to the radiation intensity of such an isotropic antenna of the same total power is called the directive gain, $G(\theta,\phi)$. Measured in dB (decibel), such a gain, G, depends upon both the directivity and efficiency of the antenna, which is given by

$$G(dB) = 10\log_{10}(G) \tag{2.354}$$

For a loss-less antenna, energy conservation requires that the gain averaged over all directions be

$$\langle G\rangle \equiv \frac{\int_{sphere} G d\Omega}{\int_{sphere} d\Omega} = 1 \tag{2.355}$$

For a loss-less antenna $\int_{sphere} G d\Omega = \int_{sphere} 1 d\Omega = 4\pi$. The gain of such an antenna depends on the angular distribution of radiation from that antenna. In an antenna having peak gain G_m, most of its power goes into a solid angle, $\Delta\Omega$, such that

$$\Delta\Omega \approx \frac{4\pi}{G_m} \tag{2.356}$$

2.7.3 Antennas for Radio Astronomy

Astronomy in radio wavebands began with the detection of cosmic noise by Karl Jansky in 1932, which appeared to be coming from the direction of

© 2015 by Taylor & Francis Group, LLC

(a) (b)

FIGURE 2.26: (a) A section of radioheliograph at the Gauribidanur Observatory, India, dedicated to observations of the solar corona in the frequency range of 30−150 MHz and (b) one of the parabolic dishes used for the Giant Metrewave Radio Telescope (GMRT) synthesis array at Narayangaon, Pune, India.

the center of the Milky Way. Unlike an optical telescope where mirrors are used to receive visible light frequencies, a radio telescope can receive waves having frequencies from a few kHz to GHz. The advantage of these frequencies is that the intervening atmosphere does not introduce any serious propagation effects and the telescopes are mainly diffraction-limited. A radio telescope can be made out of an aerial system comprising a set of dipoles (see Figure 2.26a) or a paraboloid (single-dish; see Figure 2.26b) equipped with a feed at its focus, depending on the frequency of observation ([136],[109]).

The first part of the radio receiver, referred to as the front-end, is placed right below the dipole or at the focus in the case of a dish antenna. The energy that falls on the surface of the dipole/dish forms the input to the front-end receiver. The output from the front-end receiver goes through coaxial/optical fiber cables to the main observatory building and is processed by a super-heterodyne receiver. Heterodyne is the process of down converting high frequency signal to an intermediate one by mixing it with a coherent local oscillator (LO). Such a conversion makes it easy to amplify the signal at low frequency. The intermediate frequency (IF) encompasses a range of frequencies: those greater than such an oscillator are part of the upper sideband, while those less than this are in the lower sideband. For instance, if a high frequency radio signal (RF) of frequency, ω_0, is mixed with a local oscillator of frequency ω_l ($\omega_l > \omega_0$), the upper sideband has a frequency of $\omega_l + \omega_0$ and the lower sideband has a frequency of $\omega_l - \omega_0$. The upper sideband is usually rejected by placing an appropriate filter.

In order to preserve the characteristics of the observed RF signal in the IF signal, the LO signal should, in principle, be monochromatic and the phase relationship of the LO signal at each mixer should be coherent[121]. A mixer is a device that changes the frequency of the input signal by shifting to a lower

© 2015 by Taylor & Francis Group, LLC

IF. It has two inputs, one for the signal whose frequency is to be changed; the other input is usually a sine wave generated by a tunable signal generator, the LO. Mathematically, it can be expressed as

$$e^{i2\pi ft} \times e^{-i2\pi f_{LO}t} = e^{i2\pi(f-f_{LO})t} \tag{2.357}$$

2.7.4 Waves through Ionosphere

Ionosphere is a weak natural plasma, i.e., an electrically neutral assembly of ions and electrons. The density of these electrons is a function of the height above the Earth's surface. Ionosphere plays an important role in atmospheric electricity and forms the inner edge of the magnetosphere. It can be categorized into a number of regions, for example D , E , and F regions, corresponding to peaks in the electron density. The D region lies between 60 and 95 km, the E region between 95 and 150 km, and the F region lies between 150 and 400 km. During daylight, the F region turns out to F_1 and F_2 layers, while during nighttime these two layers merge into one.

When the high frequency radio waves propagate in the ionospheric plasma, they exhibit different behaviors related to their wave frequency, oscillation frequency of the electrons in the plasma medium and the refractive index of the medium. Depending on these behaviors, the wave is refracted, reflected or attenuated by absorption from medium. Radio wave damping is due to the movements in the ionosphere of electrons and ions are caused by collisions with other particles[157]. Due to the increase of collisions, absorption increases and field strength of the radio wave decreases. As a consequence, the amplitude of the radio wave propagated in the ionosphere would decrease because of the absorption. Since the ionospheric plasma has double refractive index and generally weak conductivity in every direction, season, and local time, the ionospheric parameters, namely, conductivity, dielectric constant, and the refractive index are different in every direction and have the complex structure. To note, the refractive index is founded by using Maxwell's equations (see Section 2.4).

It is known that the refractive index of the atmosphere differs from unity; hence, there would be an additional delay in traversing the atmosphere. The accuracy during user's measurement of frequency for a ionospheric and transionospheric propagation is limited by the irregular changes of total electron content along the ray path. Besides this, the ionospheric and transionospheric signals often exhibit sudden changes of frequency during ionospheric perturbations such as that induced by solar flares or traveling ionospheric disturbances (TID). The Doppler frequency shift of radio waves reflected from the ionosphere at such times may be of the order of several Hz[45]. Also, such shift may be caused by the internal atmospheric gravity waves[46] which produces changes in the height of reflection due to the movement of the corrugated isoionic contours in the ionosphere. The ionosphere is a continually changing area, which is affected by radiation from the Sun. The radio waves on the

© 2015 by Taylor & Francis Group, LLC

short wave bands are refracted or reflected back to Earth from ionosphere. However, it is observed that signals are reduced in strength or attenuated as they pass through this area. In fact, the ionospheric absorption can be one of the major contributors to the reduction in strength of signals. In what follows, we enumerate in brief the ionospheric disturbances caused by the solar activities:

1. Solar activity: Solar activity reaches a maximum every 11 years when the Sun's magnetic field is stronger than average, which gives rise to sunspots (dark spots) on the photosphere. These spots are relatively cooler compared to the rest of the photosphere and are highly magnetized. The magnetic field in a sunspot may be 1000 times stronger than in the surrounding area, where it is about 1 G (gauss). The packed magnetic field lines provide a barrier preventing hot gas from being convected into the sunspots. Such spots, in general, occur in pairs connected by a loop. Magnetic field lines near a large group of sunspots can suddenly snap triggering a solar flare, which represents a sudden shortlived (on a time scale of about 1 hr) eruption of hot ionized gas in a localized region of the Sun. Such a flare releases large energy in the form of radiation and fast particles representing a dramatic gust in the solar wind. The solar wind is the flux particles, mainly protons and electrons together with nuclei of heavier elements in smaller numbers, which are accelerated by the high temperatures of solar corona to velocities large enough to escape from the Sun's gravitational field. When particle radiation passes into the Earth's upper atmosphere, complicated geomagnetic and ionospheric storms occur due to which radio communication gets disturbed. Bursts of solar radiation at widely different wavelengths sometimes occur during the observation of a flare in Hα and their individual characteristics differ greatly. The duration of the emission may range from less than a minute to several hours. The growth phase is usually a matter of minutes and then the flare fades slowly (tens of minutes).

   The enhancement of ionizing radiation during the solar flares produces electron density enhancements in the ionosphere. Such enhancements are more in the D region, the observed phenomena which can be detected are called the sudden ionospheric disturbances (SID) appearing in different forms in different frequency bands. These disturbances have a rapid onset phase of a few minutes followed by a relatively slow return to normal. The abrupt decrease in the signal strength of high frequency radio waves is known as shortwave fadeouts. The extent and duration of the fadeouts depend on the frequency and the length of transmission path. Low and very low frequency signals reflected from the ionosphere at both oblique and near-vertical incidence show anomalous change of phase at the time of sudden ionospheric disturbances, known as sudden phase anomaly, which is due to a real decrease in the height of the reflection level. Enhancement in the D region ionization also causes pro-

© 2015 by Taylor & Francis Group, LLC

nounced changes in the signal strength of LF−VLF atmospherics called the sudden enhancement of atmospherics. Sudden increase in the signal strength due to distant LF and VLF stations are observed as sudden enhancement of signal. A sudden frequency deviation (SFD) may be observed on an high frequency (HF) Doppler system where the received frequency increases abruptly (∼1 min) to a sharp peak. The SFDs are not sensitive to ionization enhancement of D region during solar flares, but to the higher ionospheric heights, such as, E and F regions.

2. Gravity wave perturbation effect: Atmospheric oscillations are produced by the daily heating and cooling of the atmosphere at and beneath the lower ionospheric region. These oscillatory waves propagate up into the ionosphere and the waves tend to increase in amplitude with height where the gas density decreases. The tidal winds in the dynamo region tend to carry ionization with them. This movement occurs in the presence of geomagnetic field for which the electric currents are generated. These create electric field and modify the motion and distribution of the general ion−electron distribution. As a result, the shapes of the layers are deformed. The waves of shorter period than the tides, known as gravity waves are produced by a variety of meteorological processes. They tend to increase in amplitude with height until dissipation occurs and provide a broad spectrum of irregular fluctuations at ionospheric levels ([46] and references therein). The possible sources of this gravity wave perturbations are processes associated with aurora, tidal oscillations of the atmosphere and wind systems in lower atmosphere, for instance, troposphere and stratosphere. Gravity waves may also be generated by the solar eclipse as the lunar shadow sweeps at a supersonic speed across the Earth, the cooling spot acting as a continuous source of gravity waves that build up into a bow wave.

3. Magnetic storms: Magnetic storms are not strongly flare associated but, in fact, depend magnetically upon sunspots and other solar irregularities. The solar abnormality, results in alteration of the general outflow of solar ionization at moderate energies and magnetic fields carried by that ionization. The ionization impinges on the magnetospause and alters the interaction process. Due to the storm, the ionization both of solar and of terrestrial origin is energized, with some fractions contributing to the Van Allen belts but a greater portion bombarding the polar ionosphere. The magnetic ring is produced due to the movement of energized protons and electrons. The enhanced motions are accompanied by strong electric fields which produce strong electric currents in the polar region. The energy deposited in the auroral regions is redistributed to lower latitudes; gravity waves have been the agency for this redistribution[47]. The large scale TIDs have been identified from the ionization of Joule heating and of Amperé mechanical force at the auroral zone during mag-

© 2015 by Taylor & Francis Group, LLC

netic substorms[48]. The excitation of large scale such disturbances is dependent upon the intensity of the magnetic activity[49].

Determination and dissemination of accurate measures of time has become an integral part in physical research, space navigation, global communication, and also in various astronomical observations. All these lay down stringent demands on the availability of highly accurate time and frequency signals, called the standards. These demands, to a considerable extent, have been met with the advancement of present day science and technology; the advent of atomic clocks have made available frequency standards of high accuracy. However, the ultimate precision in frequency of a standard time broadcast depends on the uncertainties in the propagation characteristics caused by the changes in various geophysical parameters. Some of these uncertainties appear to be of unpredictable nature in which many factors of solar terrestrial relationship interplay.

The time signals can be monitored anywhere on the surface of the Earth from one or more standard broadcasting transmission. The purpose of the time service is to establish local standards for measuring time and frequency. The time and frequency transmissions are either on high frequency (HF), on low frequency (LF), or on very low frequency (VLF). Most of the countries have their own service of broadcasting standard time and frequency for the regulation of the ever increasing radio communication networks. For instance in India, the standard time and frequency transmissions (5 MHz, 10 MHz, and 15 MHz) are being transmitted from National Physical Laboratory, New Delhi, using horizontal folded halfwave dipole (see Section 2.7.2.4). Investigations on the propagation characteristics of standard frequency signals are useful to explore the basic nature and origin of the changes in the ionospheric paths, which determine the stability and accuracy of the propagation path ([50],[51]).

2.8 Exercises

1. A thick slab extending from $z = -a$ to $z = a$ carries a uniform current density $\mathbf{J} = J\cdot\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the unit vector. Find the magnetic field both inside and outside the slab.

2. Find the magnetic vector potential of a finite segment of straight wire carrying a current i.

3. Derive an expression for the self-inductance per unit length of a coaxial cable having inner radius r and outer radius R when the material in the space between the conductors has the permeability μ.

4. If an alternating EMF $E = E_0 \sin \omega t$ is applied to a series LCR circuit

© 2015 by Taylor & Francis Group, LLC

(see Figure 2.15b), the resulting alternating current in the circuit is given by (steady-state) $i = i_0 \sin(\omega t \psi)$:

(a) find the current amplitude i_0 and the phase constant ψ and

(b) prove that i_0 has the maximum value (resonance) when $\omega = \omega_0$, where $\omega_0 = 1/\sqrt{LC}$ is the frequency.

5. Show that the displacement current in the dielectric of a parallel-plate capacitor is equal to the conduction current i_c in the connecting leads.

6. Derive Maxwell's equations in integral form, assuming that a charge-density distribution and current-density distribution exist in the region of interest and that $\mu = \mu_0$, $\epsilon = \epsilon_0$ for the medium under consideration.

7. Show that when integrated over a closed surface, the normal component of the Poynting vector, $\mathbf{S}$, in vacuum provides the total outward flow of energy per unit time.

8. What is meant by antenna power pattern?

9. Derive the radiation resistance of a Hertzian dipole of length:

   (a) $\frac{\lambda}{40}$

   (b) $\frac{\lambda}{60}$

   (c) $\frac{\lambda}{80}$

10. Calculate the radiated power of an antenna if a current of 10 A exists and its radiation resistance is 320 Ω.

© 2015 by Taylor & Francis Group, LLC

Chapter 3

Waves in Uniform Media

3.1 Introduction 159
3.2 Simple Harmonic Oscillation 160
3.2.1 Equation for Simple Harmonic Oscillator 162
3.2.2 The Simple Pendulum 163
3.2.3 Angular Simple Harmonic Motion 163
3.2.4 Forced Oscillation and Resonance 163
3.2.5 Resonance 164
3.3 Damped Oscillations 165
3.3.1 Damping by Friction 167
3.4 Coupled Oscillations 169
3.4.1 Superposition of Waves 169
3.4.2 Stationary Waves 169
3.4.3 Coupled Masses 169
3.5 One-Dimensional Wave Equation: D'Alembert's Solution 171
3.6 Helmholtz Equation 172
3.7 Normal Mode Eigenvalue Problem 173
3.8 Longitudinal Waves 174
3.9 Traveling Waves 176
3.9.1 Sinusoidal Traveling Waves 176
3.10 Dispersive Waves 177
3.10.1 Linear Evolution Equation 180
3.10.2 Solution of the KdV Equation 180
3.11 Solitons 181
3.11.1 Nonlinear Cubic Schrodinger Equation 184
3.11.2 Two-Soliton Solution 184
3.12 Exercises 185

3.1 Introduction

Waves and oscillations are present everywhere in nature and one encounters them in several areas of physics. An oscillation may be defined as a disturbance in a physical system that repeats in time, while a wave is a disturbance in a system which both repeats in time and is periodic spatially. In the case of a simple pendulum, an oscillation in general involves a continuous back and forth flow of both kinetic and potential energy. For example, both electromagnetic and acoustic waves are present in different physical mechanisms; however, they have many common properties. It was shown in the previous chapter the behavior of the electromagnetic waves while waves in optics will be discussed in detail in a subsequent chapter. Some of the famous texts which

© 2015 by Taylor & Francis Group, LLC

deal with waves and oscillations are by Crawford [116], French [123], Main [143], and Fitzpatrick [120].

3.2 Simple Harmonic Oscillation

Let us discuss the case of a simple harmonic oscillator from first principles. In order to do this, consider a mass **m** which is sliding over a surface which is frictionless. Assume that this mass is attached to one end of a light horizontal spring; the other end of which is fixed to a wall which is at rest. In the state of equilibrium, the mass is at rest and the spring is not stretched. In this situation, no horizontal force acts on the mass, and hence, it is stationary. When this system is slightly perturbed, the mass experiences a horizontal force (restoring), which is governed by Hooke's law which can be written mathematically as

$$f(x) = -kx \tag{3.1}$$

The constant $k > 0$ is the force constant of the spring. Since the force is a restoring force, the sign is negative. Also it is interesting to note that the magnitude of the restoring force is proportional to the displacement of the mass from its initial equilibrium position. Thus, the displacement is not very large. Applying Newton's second law of motion gives the equation for the time evolution of the system as

$$mx'' = -kx \tag{3.2}$$

Here $''$ denotes the second derivative with respect to time, i.e., $'' = d^2/dt^2$. In the literature the above equation is called the simple harmonic oscillator equation. The solution of this equation is easy to obtain and is well known in the literature. One can write a simple solution of the above equation as

$$x(t) = a\cos(\omega t - \phi) \tag{3.3}$$

The constants in the above equation are $a > 0$, $\omega > 0$, and ϕ . One can directly substitute the above solution into the harmonic equation and demonstrate that it is indeed a solution. By simple algebra one can show that ω, which may be called as the angular frequency, satisfies the relation given by

$$\omega = \sqrt{k/m} \tag{3.4}$$

a is defined as the amplitude of the oscillation. One can show that the displacement x oscillates between $-a$ and $+a$. Since the above solution is a periodic function, the motion is repetitive in time. The repetition in time is termed the period of the wave which is given by

$$T = \frac{2\pi}{\omega} \tag{3.5}$$

© 2015 by Taylor & Francis Group, LLC

The frequency of the motion can be calculated and shown to be

$$f = \frac{1}{T} = \frac{\omega}{2\pi} \tag{3.6}$$

It is easy to check that the expression $\cos\theta$ has a maximum at $\theta = n2\pi$; where n is an arbitrary integer, the maximum displacement occurs at

$$t_{max} = T\left(n + \frac{\phi}{2\pi}\right) \tag{3.7}$$

Changing the phase angle ϕ shifts the pattern of the oscillation back and forth in time. If we assume that the instantaneous displacement and velocity of the mass at $t = 0$ are x_0 and v_0, respectively, then equation (3.3) reduces to

$$x_0 = x(t=0) = a\cos\phi \tag{3.8}$$

$$v_0 = x'(t=0) = a\omega\sin\phi \tag{3.9}$$

It is well known that the cosine function is an even function so that $\cos(-\theta) = \cos\theta$, while $\sin(-\theta) = -\sin\theta$. One can show by simple calculation that

$$a = \sqrt{x_0^2 + (v_0/\omega)^2} \tag{3.10}$$

and

$$\phi = \tan^{-1}\left(\frac{v_0}{\omega x_0}\right) \tag{3.11}$$

The kinetic energy and the potential energy of the system can be calculated and written as follows:

$$\frac{1}{2}mx'^2 = \frac{1}{2}ma^2\omega^2\sin^2(\omega t - \phi) \tag{3.12}$$

$$\frac{1}{2}kx^2 = \frac{1}{2}ka^2\cos^2(\omega t - \phi) \tag{3.13}$$

so that the total energy can be written as

$$\frac{1}{2}ka^2 = \frac{1}{2}m\omega^2a^2 \tag{3.14}$$

An interesting observation at this stage is that the total energy is independent of time, i.e., a constant of the motion. Also the energy is proportional to the square of the amplitude of the oscillation. The simple harmonic motion of a mass on a spring is characterized by a continuous backward and forward flow of energy between the kinetic and potential components. When the displacement is zero, the kinetic energy is maximum, while the potential energy is minimum. When the displacement is maximum, the trend is exactly opposite.

© 2015 by Taylor & Francis Group, LLC

3.2.1 Equation for Simple Harmonic Oscillator

As in the case of the simple harmonic motion, the equation for the simple harmonic oscillator takes the following form:

$$x'' + \omega^2 x = 0 \tag{3.15}$$

where $\omega > 0$. The solution of the above equation can be written as

$$x(t) = a\cos(\omega t - \phi) \tag{3.16}$$

Here, $a > 0$ and ϕ are constants. The amplitude **a** is constant with a constant angular frequency given by ω. If we multiply equation (3.15) by x', we obtain

$$x'x'' + \omega^2 x'x = 0 \tag{3.17}$$

The above equation may be simplified to yield

$$\frac{d}{dt}\left(\frac{1}{2}x'^2\right) + \frac{d}{dt}\left(\frac{1}{2}\omega^2 x^2\right) = 0 \tag{3.18}$$

Equivalently, the above equation can be written as

$$\frac{dE}{dt} = 0 \tag{3.19}$$

where $E = (1/2)x'^2 + (1/2)\omega^2 x^2$. It is easy to prove that E is a conserved quantity, which is proportional to the overall energy of the system. The equilibrium state is characterized by $x = constant$, so the derivatives are identically zero. It is easy to check that the system will remain at rest only when $E = 0$. However, if $E > 0$, then the system will never come to a standstill and will keep moving for long periods of time. Needless to say, x reaches a maximum value when $x' = 0$. This means that

$$x_{max} = \frac{\sqrt{2E}}{\omega} \tag{3.20}$$

The solution of the simple harmonic oscillation can also be written as

$$x(t) = A\cos(\omega t) + B\sin(\omega t) \tag{3.21}$$

where $A = a\cos\phi$ and $B = a\sin\phi$. Let us consider the solution of a simple harmonic oscillator with initial conditions given by $x(0) = 1$ and $x'(0) = 0$. It is easy to check that the solution is given by

$$x_1(t) = \cos(\omega t) \tag{3.22}$$

Another solution of the simple harmonic oscillator with initial conditions interchanged is given by

$$x_2(t) = \omega^{-1}\sin(\omega t) \tag{3.23}$$

The equation being linear, a linear combination of the solutions continues to be a solution. We can write the solution as

$$x(t) = x_0\cos(\omega t) + \frac{x_0'}{\omega}\sin(\omega t) \tag{3.24}$$

© 2015 by Taylor & Francis Group, LLC

3.2.2 The Simple Pendulum

The bob of a simple pendulum undergoes nearly simple harmonic motion if the angle of twist is not very large. It is easy to calculate the period (time) of oscillation of a simple pendulum of length l which is given by

$$T = 2\pi\sqrt{l/g} \tag{3.25}$$

g in the above expression is the acceleration due to gravity.

3.2.3 Angular Simple Harmonic Motion

Assume that a disc is suspended by a thin wire. If we twist the disc from its rest position and release it, then it will oscillate about that position in an angular simple harmonic motion. Twisting the disc through an angle θ in both directions introduces a restoring torque given by

$$\Gamma = -C\theta \tag{3.26}$$

and the angular simple harmonic oscillator has a period given by

$$T = 2\pi\sqrt{I/C} \tag{3.27}$$

where I is the rotational inertia of the oscillating disc about the axis of rotation and C is the restoring torque per unit angle of twist. The angular simple harmonic motion is also referred to as a torsional pendulum in literature.

3.2.4 Forced Oscillation and Resonance

Consider a particle of mass m in the presence of an external force $F(t)\hat{i}$, in addition to the restoring force $-kx\hat{i}$ and damping force $-\beta\frac{dx}{dt}\hat{i}$. $\hat{i}$ is the unit vector. The equation of motion can be written as

$$mx'' = -kx - \beta x' + F(t) \tag{3.28}$$

If we assume that the external force is periodic, i.e., $F(t) = F\sin pt$, then we can write the above equation as follows:

$$x'' + 2bx' + \omega^2 x = f\sin(pt) \tag{3.29}$$

where $b = (\beta/2m)$, $\omega^2 = (k/m)$, and $f = (F/m)$. The equation for the forced oscillation is a nonhomogeneous linear second order differential equation, whose solution is well known in literature. For example, the general solution of the equation can be written as the linear combination of the general solution of the homogeneous equation, adding the particular solution of the nonhomogeneous part. Thus, the solution can be written as

$$x = x_1 + x_2 \tag{3.30}$$

© 2015 by Taylor & Francis Group, LLC

where x_1 is the general solution of the homogeneous equation

$$x_1'' + 2bx_1' + \omega^2 x_1 = 0 \tag{3.31}$$

and x_2 is any particular integral. The particular solution is given by

$$x_2 = \frac{f}{\sqrt{(\omega^2 - p^2) + 4b^2p^2}} \sin(pt - \alpha) \tag{3.32}$$

where

$$\tan\alpha = (2bp/(\omega^2 - p^2)), \quad 0 \leq \alpha \leq \pi \tag{3.33}$$

A simple analysis yields the result that x_1 becomes negligible within a short time, so this solution may be called the transient solution. However, after a long time when x_1 becomes negligible, the motion of the mass m is given by the expression for x_2, which is termed the steady-state solution. In the steady state, x_2 has a frequency which is equal to the frequency of the impressed force, but lags behind by a phase angle α. The vibrations or oscillations represented by x_2 are called forced vibrations or forced oscillations.

3.2.5 Resonance

The amplitude of the steady state solution x_2 is given by

$$A = \frac{f}{\sqrt{(\omega^2 - p^2)^2 + 4b^2p^2}} \tag{3.34}$$

It can be shown by simple calculation that the maximum amplitude is reached when

$$p = \omega \left[1 - \frac{2b^2}{\omega^2}\right]^{1/2} \tag{3.35}$$

where one assumes that $b^2 < \omega^2/2$. Near the frequency p of the impressed force given by the above expression, very large oscillations may set in. The phenomenon is called amplitude resonance, and the frequency is called the frequency of amplitude resonance or amplitude resonance frequency. If the damping is very small, say, $b^2 \ll \omega^2/2$, the resonance frequency, $p = \omega - b^2/\omega$, is very close to the natural frequency ω of the undamped oscillator. However, the velocity amplitude of the system becomes maximum for $p = \omega$ for any given value of b. This phenomenon is known as velocity resonance or simply resonance. One of the famous examples of resonance is the Helmholtz resonator, which is used to determine the frequency of a vibrating body with the help of the phenomena of resonance. The resonator consists of either a spherical or a cylindrical air cavity with a small neck. The dimension of the cavity is small in comparison with the wavelength of the sound to be detected. The air contained at the neck of the resonator acts like a piston alternately

© 2015 by Taylor & Francis Group, LLC

compressing and rarefying the air within the cavity of the resonator. The natural frequency of vibration of the Helmholtz resonator is given by

$$\nu = \frac{v}{2\pi}\sqrt{S/lV} \tag{3.36}$$

where v is the velocity of propagation of sound in air, l is the length of the neck of the resonator, S is the area of cross-section of the neck, and V is the volume of the resonator. The natural frequency of the resonator can be changed by changing the volume V of the resonator. When the sound wave of frequency resonant with the natural frequency of the resonator is incident on it, the resonator will produce sharp resonance. The frequency of the vibrating body is then equal to the natural frequency of the resonator.

3.3 Damped Oscillations

In the previous section we discussed physical systems which exhibit simple harmonic oscillation about a stable equilibrium state. Once the oscillation is set in, it takes a long time to die away. In practice, however, we generally encounter systems which lose energy due to reasons such as frictional or viscous heat generation while they oscillate. These systems are generally described as damped systems. In what follows, we shall discuss a damped oscillatory system.

Consider a system which possesses a mass spring system which slides over a horizontal surface. In general, the mass is subject to frictional forces and damping. This will oppose the motion, and such a force is directly proportional to its instantaneous velocity. The equation of motion of such a mass spring system whose displacement is $x(t)$ has the form given by

$$f = -kx - m\gamma x' \tag{3.37}$$

where $m > 0$ is the mass, $k > 0$ is the spring constant, and $\gamma > 0$ is assumed to be a constant which parametrizes the amount of damping. The time evolution of the system reduces to

$$x'' + \nu x' + \omega_0^2 x = 0 \tag{3.38}$$

Here, $\omega_0 = \sqrt{k/m}$ is the oscillation frequency for the undamped system. Consider a solution of the above equation in the form given by

$$x(t) = \alpha e^{-\gamma t}\cos(\omega_1 t - \phi) \tag{3.39}$$

where $\alpha > 0$, $\omega_1 > 0$, and ϕ are all constants. The above solution may be interpreted as a periodic oscillation with fixed angular frequency ω_1 and phase

© 2015 by Taylor & Francis Group, LLC

angle ϕ, whose amplitude decays exponentially in time with a form given by $\alpha(t) = \alpha \exp(-\gamma t)$. This is a generalization to the earlier solution of undamped oscillation. The damping factor introduces an exponential decay in the amplitude of the system as a function of time. Simple algebra yields the following:

$$x' = -\gamma\alpha e^{-\gamma t}\cos(\omega_1 t - \phi) - \omega_1 \alpha e^{-\gamma t}\sin(\omega_1 t - \phi) \quad (3.40)$$

$$x'' = (\gamma^2 - \omega_1^2)\alpha e^{-\gamma t}\cos(\omega_1 t - \phi) + 2\gamma\omega_1\alpha e^{-\gamma t}\sin(\omega_1 t - \phi) \quad (3.41)$$

The damped oscillator equation reduces to

$$[(\gamma^2 - \omega_1^2) - \nu\gamma + \omega_0^2]\alpha e^{-\gamma t}\cos(\omega_1 t - \phi) + [2\gamma\omega_1 - \nu\omega_1]\alpha e^{-\gamma t}\sin(\omega_1 t - \phi) = 0 \quad (3.42)$$

The above equation is valid for all times only if the coefficients of $\exp(-\gamma t)\cos(\omega_1 t - \phi)$ and the corresponding sine terms are identically zero. This means

$$(\gamma^2 - \omega_1^2) - \nu\gamma + \omega_0^2 = 0 \quad (3.43)$$

and

$$2\gamma\omega_1 - \nu\omega_1 = 0 \quad (3.44)$$

The above equations can easily be solved to yield the solutions as

$$\gamma = \nu/2 \quad (3.45)$$

$$\omega_1 = (\omega_0^2 - \nu^2/4)^{1/2} \quad (3.46)$$

The solution of the damped harmonic oscillator equation can be written as

$$x(t) = \alpha e^{-\nu t/2}\cos(\omega_1 t - \phi) \quad (3.47)$$

The above solution shows that the effect of a relatively small amount of damping, parametrized by the damping constant ν, on a system which exhibits simple harmonic motion about a stable equilibrium is to reduce the angular frequency of the oscillation from its undamped value of ω_0 to $(\omega_0^2 - \nu^2/4)^{1/2}$ and in turn causes the amplitude of the oscillation to decay exponentially in time at a rate given by $\nu/2$. The behavior of the modified oscillation which is termed the damped harmonic oscillation is depicted in Figure 3.1, for specific values of νT_0 and ϕ. For the case when the damping is sufficiently large, the system does not oscillate; instead, any motion simply decays away exponentially in time. If we assume the initial conditions, $x(0) = x_0$ and $x'(0) = \nu_0$, then substituting the conditions into the solution, yields

$$x_0 = \alpha\cos\phi \quad (3.48)$$

$$\nu_0 = -\frac{\gamma}{2}\alpha\cos\phi + \omega_1\alpha\sin\phi \quad (3.49)$$

© 2015 by Taylor & Francis Group, LLC

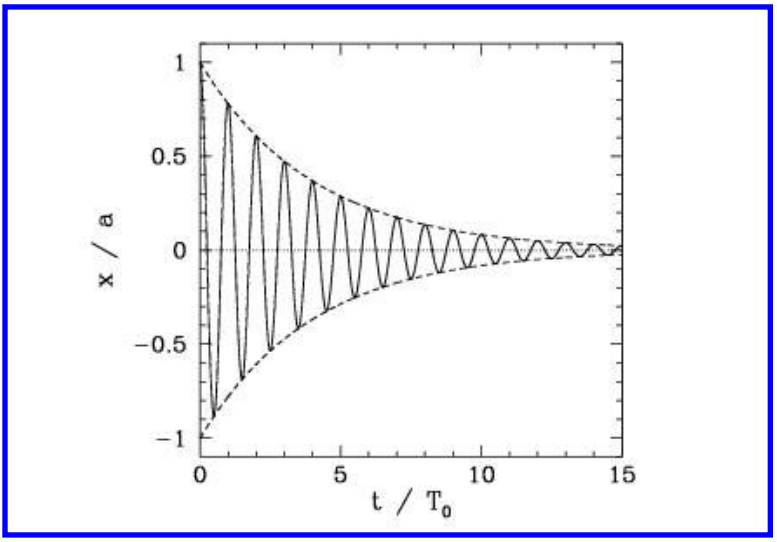


FIGURE 3.1: Damped harmonic oscillation, from Fitzpatrick[120].

giving

$$\alpha = \left[x_0^2 + \frac{(\nu_0 + \gamma x_0/2)^2}{\omega_1^2}\right]^{1/2} \tag{3.50}$$

$$\phi = \tan^{-1}\left(\frac{\nu_0 + \gamma x_0/2}{\omega_1 x_0}\right) \tag{3.51}$$

An interesting observation about the damped harmonic oscillator equation is that it is a linear differential equation in $x(t)$. Thus, a linear combination of linear solutions will also be a solution of the equation. As in the case of the undamped oscillator, the energy of the system can be calculated to yield

$$E = \frac{1}{2}mx'^2 + \frac{1}{2}kx^2 \tag{3.52}$$

The rate of change of the energy E is given by

$$\frac{dE}{dt} = -m\nu x'^2 \tag{3.53}$$

The spring force does negative work on the mass when x' and x are of the same sign and does positive work when they are of opposite sign. The spring force does no net work, while the damping force always does negative work on the mass, so that the kinetic energy of the system is reduced.

3.3.1 Damping by Friction

Consider a mass m connected to rigid supports by means of springs with a total effective spring constant k. Assume that the mass is supposed to slide

© 2015 by Taylor & Francis Group, LLC

on a rough surface and, hence, experience a constant frictional force F, which is directed in a direction opposite to the velocity. Let the initial motion be such that

$$x = a \quad x' = 0 \quad \text{at} \quad t = 0 \tag{3.54}$$

The above equation implies that the mass is pulled out at a distance a from its equilibrium position and released at $t = 0$. By Newton's law, the equation of motion can be written as

$$mx'' + kx = \pm F \tag{3.55}$$

We can transform $\omega^2 = k/m$ and $A = F/m$, so that the equation can be written as

$$x'' + \omega^2 x = +A, \quad 0 < t < \pi/\omega \tag{3.56}$$

when the force is acting in the positive x direction. If at the initial position $x = a$, $\omega^2 a > A$, then the spring force is greater than the frictional force. Consider the other scenario, wherein, the equation is

$$x'' + \omega^2 x = -A, \quad \pi/\omega < t < 2\pi/\omega \tag{3.57}$$

The equation of motion in general can be written as

$$x'' + \omega^2 x = h(t) \tag{3.58}$$

where $h(t)$ is a discontinuous function. The above equation may be solved using the Laplace transform method as given below. Assume that

$$Lx(t) = y(p) \tag{3.59}$$

$$Lh(t) = g(p) = A\left(1 - 2e^{-Tp} + 2e^{-2Tp} - 2e^{-3Tp} + \cdots\right) \tag{3.60}$$

Making use of the initial conditions, the equation reduces to

$$Lx'' = p^2 y - p^2 a \tag{3.61}$$

Substituting the expression for $Lx(t)$ and simplifying yields the following equation:

$$(p^2 + \omega^2)y = p^2 a + A\left(1 - 2e^{-Tp} + 2e^{-2Tp} - 2e^{-3Tp} + \cdots\right) \tag{3.62}$$

The above equation can be simplified to yield

$$y = \frac{p^2 a}{p^2 + \omega^2} + \frac{A}{p^2 + \omega^2}\left(1 - 2e^{-Tp} + 2e^{-2Tp} - 2e^{-3Tp} + \cdots\right) \tag{3.63}$$

From the table of Laplace transforms, one can look at the inverse of the above expression and write the solution as

$$x = a\cos\omega t + \frac{A}{\omega^2}(1 - \cos\omega t), \quad 0 < t < T \tag{3.64}$$

$$x = a\cos\omega t + \frac{A}{\omega^2}(1 - \cos\omega t) - \frac{2A}{\omega^2}[1 - \cos\omega(t - T)], \quad T < t < 2T \tag{3.65}$$

© 2015 by Taylor & Francis Group, LLC

3.4 Coupled Oscillations

3.4.1 Superposition of Waves

It is well known that the wave equation is a linear equation. If two or more waves of the same kind reach a point of the medium simultaneously, the resultant displacement ξ of the point is the vector sum of the displacements $\xi_1, \xi_2 \cdots$ of that point due to the individual waves:

$$\xi = \xi_1 + \xi_2 + \cdots \tag{3.66}$$

The principle of superposition states that when the relevant equations are linear, one can superpose any number of individual solutions to form new functions which are themselves solutions. A particular instance of this superposition, which is important in many physical problems, occurs when we add two harmonic waves going in opposite directions with the same amplitude and speed.

3.4.2 Stationary Waves

When two waves of same amplitude and frequency travel in a medium in opposite directions with the same velocity, due to superposition of two waves, there are some points of the medium which have no displacements (nodes) and there are some points which vibrate with maximum amplitudes (antinodes). The resultant wave is called a stationary wave or standing wave. The resultant wave remains confined in the region in which it is produced and is nonprogressive in character. The stationary wave is different from the progressive waves. An elaborate discussion on the above is found in Chapter 1, Sections 1.4.1 and 1.4.2.

3.4.3 Coupled Masses

Let us consider the interaction of two identical masses m of a mechanical system which are free to slide over a frictionless horizontal surface. Also assume that the masses are attached to one another and to two immovable walls, by means of three identical light horizontal springs with spring constant k. The state of the system can be specified by the displacements of the masses $x_1(t)$ and $x_2(t)$. A simple extension of the single mass to two yields the equations of motions as

$$mx_1'' = -kx_1 + k(x_2 - x_1) \tag{3.67}$$

$$mx_2'' = -k(x_2 - x_1) + k(-x_2) \tag{3.68}$$

It is assumed that a mass attached to the left end of a spring of extension x

© 2015 by Taylor & Francis Group, LLC

and spring constant k experiences a horizontal force equal to $+kx$, whereas one attached to the right has an equal and opposite force. The above equations can be rewritten as (by introducing an angular frequency)

$$x_1'' = -2\omega_0^2 x_1 + \omega_0^2 x_2 \tag{3.69}$$
$$x_2'' = \omega_0^2 x_1 - 2\omega_0^2 x_2 \tag{3.70}$$

where $\omega_0 = \sqrt{k/m}$. Let us determine a solution in which the two masses oscillate in phase at the same angular frequency, ω:

$$x_1(t) = \hat{x_1}\cos(\omega t - \phi) \tag{3.71}$$
$$x_2(t) = \hat{x_2}\cos(\omega t - \phi) \tag{3.72}$$

where $\hat{x_1}$ and $\hat{x_2}$ are constants. Equations (3.71) and (3.72) can be simplified to yield

$$-\omega^2\hat{x_1}\cos(\omega t - \phi) = (-2\omega_0^2\hat{x_1} + \omega_0^2\hat{x_2})\cos(\omega t - \phi) \tag{3.73}$$
$$-\omega^2\hat{x_2}\cos(\omega t - \phi) = (\omega_0^2\hat{x_1} - 2\omega_0^2\hat{x_2})\cos(\omega t - \phi) \tag{3.74}$$

or equivalently

$$(\hat{\omega}^2 - 2)\hat{x_1} + \hat{x_2} = 0 \tag{3.75}$$
$$\hat{x_1} + (\hat{\omega}^2 - 2)\hat{x_2} = 0 \tag{3.76}$$

where $\hat{\omega} = \omega/\omega_0$. We have, by simple algebra, converted the system of two linear differential equations into a simpler system of two coupled linear algebraic equations, which have solutions given by

$$\frac{\hat{x_1}}{\hat{x_2}} = -\frac{1}{(\hat{\omega}^2 - 2)} = -(\hat{\omega}^2 - 2) \tag{3.77}$$

The nontrivial solution is found only when

$$(\hat{\omega}^2 - 2)(\hat{\omega}^2 - 2) - 1 = 0 \tag{3.78}$$

The two algebraic equations can be written in a matrix form as follows:

$$\begin{pmatrix} \hat{\omega}^2 - 2 & 1 \\ 1 & \hat{\omega}^2 - 2 \end{pmatrix} \begin{pmatrix} \hat{x_1} \\ \hat{x_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Setting the determinant of the matrix to zero yields

$$\hat{\omega}^4 - 4\hat{\omega}^2 + 3 = (\hat{\omega}^2 - 1)(\hat{\omega}^2 - 3) = 0 \tag{3.79}$$

The roots of the above equation yield, $\hat{\omega} = 1$ or $\sqrt{3}$. These two are called the normal frequencies of the system. This is not surprising as the system possesses two degrees of freedom. In general, a dynamical system which has N degrees of freedom will possess N normal frequencies. For $\omega = \omega_0$, it can be shown that the two masses executes simple harmonic oscillation with the same amplitude and phase, which is not true for the case $\omega = \sqrt{3}\omega_0$. In this case, the amplitudes are the same with a difference in the phase. This produces a phase shift of π radians. The two distinctive patterns of motion of this system are called the normal modes of the system.

© 2015 by Taylor & Francis Group, LLC

3.5 One-Dimensional Wave Equation: D'Alembert's Solution

To begin with we shall discuss the solution of the One-dimensional wave equation, by the method developed by D'Alembert. Consider the one dimensional partial differential equation in (x,t) variables given by

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{1}{\alpha^2} \frac{\partial^2 \phi}{\partial t^2} \tag{3.80}$$

where α is a constant. Making the following transformation:

$$D_x = \frac{\partial}{\partial x}, \quad D_x^2 = \frac{\partial^2}{\partial x^2} \tag{3.81}$$

reduces the equation to a simple ordinary differential equation given by

$$\frac{d^2 y}{dt^2} = \gamma^2 y \tag{3.82}$$

Here, $\gamma = \alpha D_x$. The solution of the ordinary differential equation with γ a constant is written as

$$y = A_1 e^{\gamma t} + A_2 e^{-\gamma t} \tag{3.83}$$

where A_1 and A_2 are arbitrary constants. The formal solution of the equation can be written as

$$y = e^{\gamma t} F_1(x) + e^{-\gamma t} F_2(x) \tag{3.84}$$

Using Taylor's expansion, we can write

$$e^{hD_x} F(x) = F(x + h) \tag{3.85}$$

Letting $h = \alpha t$, the solution can be written as

$$e^{\alpha t D_x} F_1(x) = F_1(x + \alpha t) \tag{3.86}$$

For negative values of h, the other solution is

$$e^{-\alpha t D_x} F_2(x) = F_2(x - \alpha t) \tag{3.87}$$

The general solution can be written as a combination of the above solutions as

$$y(x, t) = F_1(x + \alpha t) + F_2(x - \alpha t) \tag{3.88}$$

The function $F_1(x + \alpha t)$ represents a wave of displacement of arbitrary shape traveling toward the left with a velocity equal to α, while the other function represents a wave of displacement traveling with a velocity α to the right side.

© 2015 by Taylor & Francis Group, LLC

3.6 Helmholtz Equation

In this section, using the method of separation of variables, we shall find a solution of the three-dimensional Helmholtz equation in rectangular coordinates. The differential equation is written as

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+\frac{\partial^2\phi}{\partial z^2}+k^2\phi=0 \tag{3.89}$$

Here, we shall assume that k is a constant. The separable solution has the form

$$\phi=X(x)Y(y)Z(z) \tag{3.90}$$

Substituting the above expression into the differential equation and dividing by $\phi=XYZ$, we have

$$\frac{1}{X}\frac{d^2X}{dx^2}+\frac{1}{Y}\frac{d^2Y}{dy^2}+\frac{1}{Z}\frac{d^2Z}{dz^2}+k^2=0 \tag{3.91}$$

Since $X(x)$ is a function of only x, it is evident that $(1/X)d^2X/dx^2$ will be independent of y and z. On the other hand, we see that the quantity $(1/X)d^2X/dx^2$ is equal to the quantity

$$-k^2-\frac{1}{Y}\frac{d^2Y}{dy^2}-\frac{1}{Z}\frac{d^2Z}{dz^2} \tag{3.92}$$

which is a function of y and z only. The above statements imply that

$$\frac{1}{X}\frac{d^2X}{dx^2}=-k_1^2 \tag{3.93}$$

In a similar way

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -k_2^2 \tag{3.94}$$

$$\frac{1}{Z}\frac{d^2Z}{dz^2} = -k_3^2 \tag{3.95}$$

Substituting the above expressions leads to the relation

$$k_1^2+k_2^2+k_3^2=k^2 \tag{3.96}$$

The above differential equations are easy to solve and their solutions can be written as

$$X = A_1\exp(ik_1x)+B_1\exp(-ik_1x) \tag{3.97}$$

$$X = A_2\exp(ik_2x)+B_2\exp(-ik_2x) \tag{3.98}$$

$$X = A_3\exp(ik_1x)+B_3\exp(-ik_3x) \tag{3.99}$$

© 2015 by Taylor & Francis Group, LLC

Multiplying the above three equations, one gets the solution for the original Helmholtz equation:

$$\phi(\bar{r}) = Ae^{i\bar{k}\cdot\bar{r}} \tag{3.100}$$

with the condition that $\bar{k}^2 = k^2$.

3.7 Normal Mode Eigenvalue Problem

In this section, we shall study the normal modes of the wave equation by reducing it into an eigenvalue problem and discuss its solution. Consider the diffusion equation:

$$\boldsymbol{\nabla}^2\psi(\bar{r},t) = \frac{1}{\kappa}\frac{\partial\psi(\bar{r},t)}{\partial t} \tag{3.101}$$

The above equation will be solved in a given volume V, subject to the boundary condition and an initial condition $\psi(\bar{r},0)$ given throughout the volume V and $\psi(\bar{r},t)$ throughout the surface S. Let us assume the following expression:

$$\psi(\bar{r},t) = \psi'(\bar{r}) + \psi''(\bar{r},t) \tag{3.102}$$

so that $\psi''(\bar{r},t)$ is determined by the equation

$$\boldsymbol{\nabla}^2\psi''(\bar{r},t) = \frac{1}{\kappa}\frac{\partial\psi''(\bar{r},t)}{\partial t} \tag{3.103}$$

with a homogeneous boundary condition:

$$\psi''(\bar{r},t) = 0, \quad \text{for} \quad \bar{\text{r}} \quad \text{on} \quad S \tag{3.104}$$

We separate the time dependence and write the solution for ψ as

$$\psi''(\bar{r},t) = e^{-\kappa k^2 t}u(\bar{r}) \tag{3.105}$$

$$\boldsymbol{\nabla}^2u(\bar{r}) + k^2u(\bar{r}) = 0 \tag{3.106}$$

The homogeneous boundary condition for u can be written as

$$u(\bar{r}) = 0 \quad \text{for} \quad \bar{r} \quad \text{on} \quad S \tag{3.107}$$

The above Helmholtz equation with the boundary condition forms an eigenvalue problem, which is the Sturm−Liouville problem. There are an infinite set of solutions, i.e., eigenfunctions $u_n(\bar{r})$ and eigenvalues k_n^2, satisfying the equation

$$\boldsymbol{\nabla}^2u_n(\bar{r}) + k^2u_n(\bar{r}) = 0 \tag{3.108}$$

with the boundary condition

$$u_n(\bar{r}) = 0, \quad \text{for} \quad \bar{r} \quad \text{on} \quad S \tag{3.109}$$

© 2015 by Taylor & Francis Group, LLC

Using Green's theorem, one can easily show that these eigenfunctions are orthogonal to each other. These eigenfunctions form a complete set and we can expand any arbitrary function $f(\bar{r})$ as

$$f(\bar{r}) = \sum_n c_n u_n(\bar{r}) \tag{3.110}$$

Using the orthogonality condition, one can determine c_n as follows:

$$c_n = \frac{\int_V d^3x u_n(\bar{r}) f(r)}{\int_V d^3x [u_n(r)]^2} \tag{3.111}$$

Returning to the diffusion equation, we see that we have produced an infinite set of solutions satisfying the boundary condition (given apriori):

$$\psi''(r,t) = \sum_n c_n e^{-\kappa k_n^2 t} u_n(\bar{r}) \tag{3.112}$$

The coefficients c_n can now be written in terms of ψ as follows:

$$c_n = \frac{\int_V d^3x u_n(\bar{r}) \psi''(\bar{r},0)}{\int_V d^3x [u_n(\bar{r})]^2} \tag{3.113}$$

This concludes the discussion on the solution of the eigenfunctions of the diffusion equation by converting it into an eigenvalue problem.

3.8 Longitudinal Waves

In this section, we shall discuss waves which are longitudinal in nature. Waves in general can be classified as transverse and longitudinal. A transverse disturbance, traveling along a string, makes each part of the string move sideways. However, longitudinal waves disturb the system by producing actual motion along the direction in which the waves travel. In what follows, we shall discuss longitudinal disturbances of a stretched string or wire. These results are also valid for rigid objects such as cylindrical rods, assuming that they are not thick. Waves in liquids and gases will be discussed in Chapter 4.

Consider small distortions involving local compressions and extensions of the string along its own axis, which we will assume is the z axis. As a result of such a distortion, one segment between the parallel planes z and $z + \triangle z$, from the equilibrium will get distorted into another segment, which also lies between the two planes. Assume the left boundary plane is displaced from z to a new position $z + \phi$, and that the other boundary on the right moves from $z + \triangle z$ to $(z + \triangle z) + (\phi + \triangle\phi)$ in such a way that

$$|\triangle\phi| \ll |\triangle z| \tag{3.114}$$

© 2015 by Taylor & Francis Group, LLC

The deformation is such that the segment is both stretched and shifted. Its cross-section may also get changed slightly. However, we may neglect this effect. Let us denote the magnitude of the new stress force stretching by F and by $\triangle F$ with the excess force accelerating the whole segment to the right. The corresponding strain due to the increase in the length segment is given by $\triangle\phi/\triangle z$. This is related to the stress F/a by

$$\frac{F}{a} = E\left(\frac{\triangle\phi}{\triangle z}\right) \tag{3.115}$$

where a is the cross-sectional area and E, Young's modulus. If we now assume that the segment is sufficiently thin, then

$$F = aE\left(\frac{\partial\phi}{\partial z}\right) \tag{3.116}$$

The additional force which is responsible for accelerating the segment is given by

$$\triangle F = aE\left(\frac{\partial^2\phi}{\partial z^2}\right)\triangle z \tag{3.117}$$

The mass of the segment may be approximated by $\rho a\triangle z$, where ρ is the density, and the acceleration of the mass center is approximately equal to $\frac{\partial^2\phi}{\partial t^2}$. Using Newton's second law which relates force to acceleration, we have

$$\rho a\left(\frac{\partial^2\phi}{\partial t^2}\right)\triangle z \approx aE\left(\frac{\partial^2\phi}{\partial z^2}\right)\triangle z \tag{3.118}$$

Canceling $\triangle z$ on either side of the above equation, leads to the well-known wave equation given by

$$\frac{\partial^2\phi}{\partial t^2} \approx \frac{E}{\rho}\left(\frac{\partial^2\phi}{\partial z^2}\right) \tag{3.119}$$

Letting $c = (E/\rho)^{1/2}$, gives the speed of the longitudinal traveling waves, on a string or a wire whose density and Young's module are known specifically. An important comment at this stage is that the phase speeds of the longitudinal traveling waves are in general larger than the transverse wave speeds. In order to produce such a transverse wave with a phase speed equal to that of a traveling wave, one should apply a stress T/a, equal to that of Young's modulus of the material. Ordinary strings may not be able to stand such a stress. A word of caution: The analysis mentioned above is strictly not valid for infinite planes, as the discussion involves Young's modulus, which is applicable to situations in which the material is free to contract sideways as it is stretched.

© 2015 by Taylor & Francis Group, LLC

3.9 Traveling Waves

The wave equation is a partial differential equation which is linear and homogeneous, with constant coefficients. Let us consider the wave equation in one dimension given by

$$\frac{\partial^2 \phi}{\partial t^2} = c^2 \left(\frac{\partial^2 \phi}{\partial z^2} \right) \tag{3.120}$$

The above equation has two variables, z and t. We shall look for solutions for the above equation in the form given by

$$\phi = C \exp[i(\omega t - kz)] + C^* \exp[-i(\omega t - kz)] \tag{3.121}$$

The plus or minus sign in front of the term k is purely a matter of convenience. This does not change the nature of the solution. The most general solution can have both possibilities. We shall discuss the case when k is either positive or negative. A solution of this equation will represent a motion in which every point on the string vibrates harmonically with an angular frequency ω, which we may assume without loss of generality to be positive. Substituting the form of the solution into the wave equation results in a relation between ω and k given by

$$\omega^2 = c^2 k^2 \tag{3.122}$$

In other words, $\omega = \pm ck$. The above relation is called the dispersion relation for the traveling waves. As was already mentioned in the introductory chapter, the dispersion relation is a relation between the frequency and the wavenumber. The phase speed defined as ω/k is a constant in this case. This is under the assumption that c is assumed to be constant. As the phase speed is a constant, this wave will be termed nondispersive. We shall generalize the solution and talk about sinusoidal traveling waves in the next section.

3.9.1 Sinusoidal Traveling Waves

The form of the solution for the wave equation can also be written as

$$\phi = A \cos(\omega t - kz + \psi) \tag{3.123}$$

It is easy to realize that each point on the string vibrates harmonically, and all points have the same frequency. The amplitudes of these waves are also the same. The phase speed is constant, with the exception that it increases or decreases linearly with distance z along the string. This feature is different from the one we discussed in the previous section. At any given time (fixed t), the shape of the string is sinusoidal, and repeats itself at a distance $2\pi/|k|$. However, if we concentrate on a particular point (fixed z), allowing the time to increase from t to $t + \triangle t$, we will realize that the displacement changes from

© 2015 by Taylor & Francis Group, LLC

$\phi(z,t)$ to $\phi(z,t+\triangle t)$. The displacements in the other points may also change. However, it is possible that at one of the points, say, $z+\triangle z$, the displacement $\phi(z+\triangle z, t+\triangle t)$ may be equal to the original displacement $\phi(z,t)$ at some point z. This is possible if the corresponding phase angles are equal. That is to say,

$$\omega t - kz = \omega(t+\triangle t) - k(z+\triangle z) \tag{3.124}$$

$$\omega\triangle t - k\triangle z = 0 \tag{3.125}$$

The above expression implies $\triangle z/\triangle t = \omega/k$. For sufficiently small values of $\triangle z$ and $\triangle t$, we can write

$$\frac{dz}{dt} = \frac{\omega}{k} = v_\phi \tag{3.126}$$

The above argument is true for all values of z. This means that the whole sinusoidal profile moves along to either the right or left side of the string with a speed given by v_ϕ, which is nothing but the phase velocity, as it denotes the velocity at a single point in the phase angle. The profile moves in the positive z direction if k is positive, and vice versa. The motion described above is termed a sinusoidal traveling wave. The actual value of the phase velocity v_ϕ for any particular system depends on how ω and k are related to each other in the system. The distance after which the wave pattern will repeat itself is the wavelength given by

$$\lambda = 2\pi/|k| \tag{3.127}$$

The relation for the wavelength in terms of the frequency is given by $c = \nu\lambda$, where ν is the frequency of the system. In the next section we shall discuss wave propagation for dispersive systems.

3.10 Dispersive Waves

In the previous sections, we have been discussing wave propagation wherein the dispersion relation between the frequency and the wavenumber was very simple. Here, in this section, we shall include effects where the waves become dispersive. The discussion and analysis is far from being trivial. However, we shall restrict ourselves to simpler systems. Consider a one-dimensional wave described by

$$\phi(x,t) = \int_{-\infty}^{+\infty} C(k)\cos(kx-\omega t)dk \tag{3.128}$$

The above expression is a linear superposition of cosine waves with a large range of wavenumbers, all of which are traveling in the positive x direction. The angular frequency ω of each of these waves is related to the wavenumber

© 2015 by Taylor & Francis Group, LLC

k, via a relation which is termed the dispersion relation, which in general can be written as

$$\omega = \omega(k) \tag{3.129}$$

The above relation can be derived from the equation of motion. We have so far discussed linear dispersion relation given by

$$\omega = k\nu \tag{3.130}$$

Here, ν is a constant. The above relation implies that the waves will have the same phase velocity given by ν. Substituting this relation into the expression for ϕ, we have

$$\phi(x,t) = \int_{-\infty}^{+\infty} C(k)\cos[k(x-\nu t)]dk \tag{3.131}$$

which describes a wave propagating in the positive x direction, at a fixed speed ν, without changing the shape. There are many types of sinusoidal waves whose dispersion relations are nonlinear. For example, the dispersion relation for sinusoidal electromagnetic waves propagating through magnetized plasma (see Chapter 9) is given by

$$\omega = \sqrt{k^2c^2 + \omega_p^2} \tag{3.132}$$

where c is the speed of light in vacuum, and ω_p is a constant, known as the plasma frequency, which depends on the properties of the plasma. Also the dispersion relation for sinusoidal surface waves in deep water is (see Chapter 4)

$$\omega = \sqrt{gk + \frac{T}{\rho}k^3} \tag{3.133}$$

where g is the acceleration due to gravity, T is the surface tension, and ρ is the density. Sinusoidal waves which satisfy nonlinear dispersion relations mentioned above are known as dispersive waves, as opposed to waves which satisfy linear dispersion relations. An important question that needs to be addressed is the behavior of a wave made up of a linear superposition of dispersive sinusoidal waves evolved in time. If we assume that $C(k)$ has a form like

$$C(k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left[-\frac{(k-k_0)^2}{2\sigma_k^2}\right] \tag{3.134}$$

this means that $C(k)$ is a Gaussian, with a characteristic width σ_k, centered on the wavenumber $k = k_0$. It is well known from the properties of a Gaussian function that $C(k)$ will be negligible for those wavenumbers such that $|k-k_0| \geq 3\sigma_k$. This means that the significant contributions to the integral will appear as

$$\phi(x,t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left[-\frac{(k-k_0)^2}{2\sigma_k^2}\right]\cos(kx-\omega t)dk \tag{3.135}$$

© 2015 by Taylor & Francis Group, LLC

which is centered around a small region in the k space around $k = k_0$. We shall invoke the Taylor expansion of the dispersion relation $\omega = \omega(k)$ around the point $k = k_0$. If we neglect the higher order terms in the expansion (second order and above), we get

$$\omega \approx \omega(k_0) + (k - k_0)\frac{d\omega(k_0)}{dk} \tag{3.136}$$

It follows that

$$kx - \omega t \approx k_0 x - \omega_0 t + (k - k_0)(x - v_g t) \tag{3.137}$$

where $\omega_0 = \omega(k_0)$ and $v_g = d\omega(k_0)/dk$ is a constant. If we assume that σ_k is sufficiently small, we can neglect second order terms in the above expansion, and the contribution to the dispersive wave will be

$$\begin{aligned}\phi(x,t) \quad \approx \quad & \frac{\cos(k_0 x - \omega_0 t)}{\sqrt{2\pi\sigma_k^2}} \int_{-\infty}^{+\infty} \exp\left[-\frac{(k-k_0^2)}{2\sigma_k^2}\right] \cos[(k-k_0)(x-v_g t)]dk \\ & -\frac{\sin(k_0 x - \omega_0 t)}{\sqrt{2\pi\sigma_k^2}} \int_{-\infty}^{+\infty} \exp\left[-\frac{(k-k_0^2)}{2\sigma_k^2}\right] \sin[(k-k_0)(x-v_g t)]dk\end{aligned} \tag{3.138}$$

The second integral is zero by virtue of its being a symmetric function. Using the fact that

$$\frac{1}{\sqrt{2\pi\sigma_k^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{k^2}{2\sigma_k^2}\right) \cos(kx)dk = \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \tag{3.139}$$

where $\sigma_x = 1/\sigma_k$, the expression for $\phi(x,t)$ can be written as

$$\phi(x,t) \approx \exp\left[-\frac{(x - v_g t)^2}{2\sigma_x^2}\right] \cos(k_0 x - \omega_0 t) \tag{3.140}$$

This is clearly the equation of a wave pulse, of wavenumber k_0 and angular frequency ω_0, with a Gaussian envelope, of characteristic width σ_x, whose peak (which is located by setting the argument of the exponential to zero) has the equation of motion $x = v_g t$. In other words, the pulse peak, and hence the pulse itself propagates at the velocity v_g, which is known as the group velocity. Of course, in the case of nondispersive waves, the group velocity is the same as the phase velocity (since if $\omega = kv$ then $\omega/k = d\omega/dk = v$). However, for the case of dispersive waves, the two velocities are, in general, different. The spatial extent of the wave in the real space is given by

$$\triangle x \approx (\triangle x)_0 + \frac{d^2\omega(k_0)}{dk^2}\frac{t}{(\triangle x)_0} \tag{3.141}$$

We, thus, conclude that the spatial extent of the wave grows linearly in time,

© 2015 by Taylor & Francis Group, LLC

at a rate proportional to the second derivative of the dispersion relation with respect to k (evaluated at the waves' central wavenumber). This effect is known as wave dispersion. In summary, a wave made up of a linear superposition of dispersive sinusoidal waves, with a range of different wavenumbers, propagates at the group velocity and also gradually disperses as time progresses.

3.10.1 Linear Evolution Equation

It is well known that initial and boundary value problems associated with nonlinear partial differential equations are very difficult to handle in a general way. However, some specific problems can be tackled from time to time by methods specifically suited to the individual problems. Nonlinear waves, are governed by nonlinear partial differential equations. To begin with we shall consider the linearized version of a nonlinear equation given by

$$u_t + cu_x + Ku_{xxx} = 0, \quad c, \quad K \quad \text{constant}, \quad K > 0 \tag{3.142}$$

For simplicity, consider the case when $K > 0$. It is interesting to note that the nonlinear equation given by

$$u_t + uu_x + Ku_{xxx} = 0, \quad K < 0 \tag{3.143}$$

can be transformed into

$$u_t + uu_x + (-K)u_{xxx} = 0, \quad , -K > 0 \tag{3.144}$$

by the transformation $u \to -u, \quad x \to -x, \quad t \to t$. Applying the Fourier method to the linear equation, we get the following dispersion relation

$$\omega = ck - Kk^3 \tag{3.145}$$

The ω is a real function of k and we can define the phase velocity and the group velocity in the usual manner as

$$V_p = \omega/k = c - Kk^2 \tag{3.146}$$

and

$$V_g = \omega'(k) = c - 3Kk^2 \tag{3.147}$$

The fact that $V_p \neq V_g$ implies that this is a dispersive system.

3.10.2 Solution of the KdV Equation

Let us consider the linearized form of the KdV equation:

$$u_t + Ku_{xxx} = 0 \tag{3.148}$$

for which the dispersion relation is

$$\omega = -Kk^3 \tag{3.149}$$

© 2015 by Taylor & Francis Group, LLC

Again, since $\omega'(k) \neq 0$, the linearized K dV equation is dispersive. Let us seek a steady solution of the K dV equation in the form

$$u(x,t) = u(\xi), \quad \xi = x - ct \tag{3.150}$$

Substituting the above relation into the K dV equation results in

$$-cu_\xi + uu_\xi + Ku_{\xi\xi\xi} = 0 \tag{3.151}$$

Integrating the above equation yields

$$-cu + \frac{1}{2}u^2 + Ku_{\xi\xi} = A \tag{3.152}$$

Multiplying the above equation by u_ξ and integrating, we get

$$-\frac{1}{2}cu^2 + \frac{1}{6}u^3 + \frac{1}{2}Ku_\xi^2 = Au + B \tag{3.153}$$

or

$$3Ku_\xi^2 = -u^3 + 3cu^2 + 6Au + 6B = f(u) \tag{3.154}$$

If we write it in the form

$$\frac{1}{2}u_\xi^2 + \frac{1}{6K}[-f(u)] = 0 \tag{3.155}$$

and interpret u and ξ as space and time coordinates, respectively, we may interpret the above equation as the energy equation for the motion of a particle with unit mass under the action of the potential $(1/6K)f(u)$, or as an equation for an anharmonic oscillator under certain periodic motion oscillating between two consecutive real zeros of $f(u)$, where $f(u) \geq 0$. This provides an important clue for studying the evolution equation. Now $f(u)$ is cubic and has three zeros. $u = c$ is a trivial solution of the above equation (Equation 3.155). Some important properties of these solutions are (1) the product of the amplitude of the pulse c times the square of the width $(1/\sqrt{c})$ is a constant and (2) its velocity in the laboratory frame equals $c_0(1 + c)$ and is amplitude dependent where c_0 is the linear velocity of the system. A soliton differs from a solitary waves in its ability to survive a collision with another soliton. A large amplitude soliton can catch up with a smaller one and just pass through it. The discussion on the other nontrivial solutions of the equation is highly technical and is skipped for brevity.

3.11 Solitons

In this section, we shall determine the solution of the KdV equation with the nonlinearity included. Consider the equation

$$u_t + uu_x + Ku_{xxx} = 0 \tag{3.156}$$

© 2015 by Taylor & Francis Group, LLC

Using the transformation $x \to K^{1/3}x, \quad u \to -6K^{1/3}u, \quad t \to t$, the above equation can be reduced to

$$u_t - 6uu_x + u_{xxx} = 0 \tag{3.157}$$

The initial condition is assumed to be $u(x,0) = u_0(x)$ which is bounded and continuously differentiable. Also $u(x,t)$ and its derivatives tend to zero as $|x| \to \infty$. The solitary wave solution can be written (under the above conditions) as

$$u(x,t) = -\frac{a^2}{2}\sec h^2\left[\frac{a}{2}(x - a^2t)\right] \tag{3.158}$$

where a is a function of K.

A soliton is a special solution of solitary waves where the interaction between two or more solitary waves does not change the shape of the wave. The analysis dwells upon associating the K dV equation with the steady one-dimensional Schrodinger equation. Consider the time-dependent one-dimensional Schrodinger equation given by

$$\phi_{xx} + [\lambda - u(x)]\phi = 0 \tag{3.159}$$

where $u(x)$ is the potential, λ is the energy eigenvalue, and ϕ is the wave function.

Some of the properties of the eigenvalues are as follows:

1. The eigenvalues λ may be discrete or continuous or both in a given system.

2. The discrete eigenvalues are negative and correspond to the stable states of the finite motion of the particle, where by finite motion, we mean the motion which is confined to finite bounded space. We shall denote the discrete eigenvalues by $\lambda = -\kappa_1^2, -\kappa_2^2, -\kappa_3^2, \cdots, -\kappa_m^2$, where $k_i > 0$,

3. The continuous eigenvalues correspond to infinite motion in which the particle reaches infinity. At sufficiently large distances the potential field $u(x)$ may be neglected and the particle may be regarded as free. The energy of a free particle is positive. Thus, the continuous eigenvalues are positive, say, $\lambda = k^2, \quad k > 0$.

4. None of the discrete eigenvalues is degenerate, i.e., for each discrete eigenvalue, there is one and only one eigenfunction.

Consider a potential $u(x) \to 0$ as $|x| \to \infty$. Then when $|x| \to \infty$, the Schrodinger equation assumes the following asymptotic form:

$$\phi_{xx} + \lambda\phi = 0 \tag{3.160}$$

For discrete eigenvalues, it takes the form

$$\phi_{xx} - \kappa^2\phi = 0 \tag{3.161}$$

© 2015 by Taylor & Francis Group, LLC

and its two independent solutions are

$$\phi = c_{\pm} \exp(\pm \kappa x) \tag{3.162}$$

When $x \to \infty$, the admissible solution is $\exp(-\kappa x)$, and when $x \to -\infty$, the admissible solution is $\exp(\kappa x)$. For continuous eigenvalues, the Schrodinger equation takes the form

$$\phi_{xx} + k^2 \phi = 0 \tag{3.163}$$

whose solution can be written down immediately in terms of the exponential functions. The K dV equation given by

$$u_t - 6uu_x + u_{xxx} = 0 \tag{3.164}$$

can be written in the conservation form as

$$T_t + X_x = 0 \tag{3.165}$$

where $T = u$, and $X = -3u^2 + u_{xx}$. If we assume that u is periodic in x or that u and its derivatives vanish sufficiently rapidly at the two infinities $x = \pm\infty$, integrating the conservation law, we get

$$\frac{\partial}{\partial t} \int T dx = 0, \quad \text{or} \quad I = \int T dx = \text{independent} \quad \text{of} \quad \text{time} \tag{3.166}$$

Let us set $u = v^2 + v_x$ in the KdV equation so that we have

$$2v(v_t - 6v^2 v_x + v_{xxx}) + (v_t - 6v^2 v_x + v_{xxx})_x = 0 \tag{3.167}$$

For a given function u, the relation $u = v^2 + v_x$ becomes a Riccati equation which can be linearized by the well-known transformation $v = \phi_x/\phi$. More discussion about the Riccati equation is found in later chapters.

Consider a special form of the potential for the Schrodinger equation given by $u_0(x) = -2 \sec h^2 x$. The Schrodinger equation reduces to

$$\phi_{xx} + (2 \sec h^2 x + \lambda)\phi = 0 \tag{3.168}$$

If we substitute $\phi = \omega \sec h^2 x$, then the above equation reduces to

$$\omega_{xx} - 2s(\tanh x)\omega_x + \omega[(2 - s - s^2) \sec h^2 x + \lambda + s^2] = 0 \tag{3.169}$$

We choose s such the coefficient of ω is independent of x. The values of s are 1 and -2, respectively. Choosing $s = 1$, we have

$$\omega_{xx} - 2(\tanh x)\omega_x + (\lambda + 1) = 0 \tag{3.170}$$

Setting $\xi = \sinh^2 x$, the equation is transformed to

$$\xi(1 + \xi)\omega_{\xi\xi} + (1/2)\omega_\xi + (1/4)(1 - \kappa^2)\omega = 0 \tag{3.171}$$

Setting $\omega = (1 + \xi)^{1/2}\phi$ and $\lambda = -\kappa^2$, the solution can be derived (details are skipped) as

$$\phi = -2 \sec h^2(x - 4t) \tag{3.172}$$

which is nothing but the solitary wave solution. Thus, in this analysis we have brought out the relationship between the solitary wave solution of the KdV equation and the one-dimensional Schrodinger equation.

© 2015 by Taylor & Francis Group, LLC

3.11.1 Nonlinear Cubic Schrodinger Equation

In this section, we shall briefly describe a solution of the nonlinear cubic Schrodinger equation which has a form given by

$$iu_t + u_{xx} + a_1 u|u|^2 - a_2 u|u|^3 = 0 \tag{3.173}$$

where a_1 and a_2 are real constants.

This equation has several applications, when one considers unidirectional propagation of wave packets in a dispersive energy conserving medium, such as nonlinear pulses on an optical fiber, two-dimensional self-focusing of a plane wave, propagation of a heat pulse in a solid, one-dimensional self-modulation of a monochromatic wave, and Langmuir waves in plasmas, to name a few. Assume that $a_1 = \nu$ and introduce a moving coordinate $X = x - Ut$. Introduce the variable given by

$$u = e^{irx - ist} v(X) \tag{3.174}$$

where r and s are constants. On substitution, the ordinary differential equation for v can be written as

$$v'' + i(2r - U)v' + (s - r^2)v + \nu|v|^2 v = 0 \tag{3.175}$$

If we now choose $r = U/2$ and $s = U^2/4 - \alpha$, we can eliminate the term in v'. The resulting equation is

$$v'' - \alpha v + \nu v^3 = 0 \tag{3.176}$$

The above equation in the literature is often referred to as the cnoidal wave equation. Integrating once, we get

$$v'^2 = A + \alpha v^2 - \frac{\nu}{2} v^4 \tag{3.177}$$

which can be solved in terms of elliptic functions. The limiting case of a solitary wave solution for the above equation results when $\nu > 0$, $A = 0$, and $\alpha > 0$. The resulting solution is given by

$$v = \left(\frac{2\alpha}{\nu}\right)^{1/2} \sec h \left[\alpha^{1/2}(x - Ut)\right] \tag{3.178}$$

3.11.2 Two-Soliton Solution

The starting point for this discussion is the Schrodinger equation given by

$$\frac{\partial^2 \phi}{\partial x^2} + (\lambda - u)\phi = 0 \tag{3.179}$$

so that the solution can be written as

$$\phi_1 = \epsilon^{i\sqrt{\lambda}x}[2\sqrt{\lambda} + ia(x)] \tag{3.180}$$

$$\phi_2 = \epsilon^{i\sqrt{\lambda}x}[4\lambda + 2i\sqrt{\lambda}a(x) + b(x)] \tag{3.181}$$

© 2015 by Taylor & Francis Group, LLC

where $a(x)$ and $b(x)$ are functions of x. Substituting the solution ϕ_2 into the Schrodinger equation and defining $a(x) = (d(ln\omega(x))/dx)$, the solution for $\phi(x,t)$ after algebra can be written as

$$\phi(x,t) = -12\left[\frac{4\cosh(2x-8t)+\cosh(4x064t)+3}{[3\cosh(x-28t)+\cosh(3x-36t)]^2}\right] \tag{3.182}$$

At $t=0$, the two-solitons overlap and $\phi(x,0) = -6\sec h^2 x$. Introducing additional effects of transverse inhomogeneity into the K dV equation, results in the famous Kadomtsev−Petviashvili equation given by

$$\frac{\partial}{\partial x}\left[\frac{\partial\phi}{\partial t}+\phi\frac{\partial\phi}{\partial x}+\frac{1}{2}\frac{\partial^3\phi}{\partial x^3}\right]+\frac{1}{2}\frac{\partial^2\phi}{\partial y^2}=0 \tag{3.183}$$

The above equation is a nonlinear evolution equation which admits soliton solutions given by

$$\phi = 6k^2\sec h^2[kx+ly-\omega t] \tag{3.184}$$

provided that

$$\omega = 2k^3+\frac{l^2}{2k} = \omega(\bar{K}) \tag{3.185}$$

where $\bar{K} = k\bar{\phi}_x + l\bar{\phi}_y$ is satisfied. It is important to realize that the dispersion relation mentioned above is not the relation obtained by linearizing the K P equation. Additional properties of the solitons can be found in the book by Hirose and Lonngren [132]. More discussion on the KdV equation and solitary waves will be dealt with in detail when studying their properties in hydrodynamic and plasma flows.

3.12 Exercises

1. Assume that a point executes a simple harmonic motion with a period of π seconds. When it passes through the center of its path, its velocity is 0.2 m/sec. What is the velocity when it is at a distance 0.05 m from the mean position?

2. A mass of 5 g vibrates through 1 mm on each side of the middle point of its path and makes 600 complete vibrations per second. Assuming simple harmonic motion, calculate the maximum force acting on the particle.

3. Let ω be the frequency of two simple harmonic motions $x_1 = a_1\sin\omega t, x_2 = a_2\sin(\omega t+\psi)$ acting on a particle along the x axis simultaneously. Find the resultant motion.

© 2015 by Taylor & Francis Group, LLC

4. A mass equal to 3 g moves along the x axis attracted toward origin by a force whose magnitude is 12. If the particle is at rest at $x = 10$ and is subject to a damping force equal to 12 times the instantaneous speed, calculate the position and velocity at any time.

5. Calculate the velocity and the maximum amplitude of a mass m when it is subject to steady-state forced vibration whose displacement is $A\sin(pt - \alpha)$.

6. Calculate the general expression for the standing wave solution of the wave equation
$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

7. Write the Helmholtz equation in cylindrical coordinates (r, θ, z) and sketch the solution by separation of variables.

8. Derive the dispersion relation for the linearized KdV equation.

© 2015 by Taylor & Francis Group, LLC

Chapter 4

Hydrodynamic Waves

4.1 Introduction 187
4.2 Basic Equations 188
4.2.1 Equations in a Rotating Frame 190
4.3 Small-Amplitude Waves 190
4.3.1 An Application in Geophysics 192
4.4 Gravity Waves 192
4.4.1 Waves in a Steady Stream 195
4.5 Linear Capillary and Gravity Waves 195
4.5.1 One-Dimensional Capillary – Gravity Waves 198
4.6 Surface Waves Generated by a Local Disturbance in the Field 199
4.7 Klein–Gordon Equation 202
4.8 Shallow Water Waves 203
4.8.1 Long Waves in Shallow Water 205
4.9 Boussinesq Equation 208
4.10 Finite Amplitude Shallow Water Waves (Nonlinear Aspects) 210
4.11 Plane Waves in a Layer of Constant Depth 212
4.12 Poincaré and Kelvin Waves 214
4.13 Lamb and Rayleigh Waves 218
4.14 Inertial Waves 221
4.14.1 Axisymmetric Waves 223
4.15 Rossby Waves 223
4.16 Forced Stationary Waves in the Atmosphere 225
4.17 Solitary Waves – KdV Equation 227
4.17.1 KdV Equation 231
4.17.2 Solitary Waves 231
4.18 Exercises 232

4.1 Introduction

The introductory chapter has basic definitions of a wave, basic equations of motion, different conservation laws, the concept of phase and group velocities, and the resulting notion of dispersion relation. Chapters 2 and 3 deal with waves in electromagnetic fields and the different modes in an uniform media. In this chapter, for the first time, we shall discuss waves and oscillations in fluids (both liquids and gases). We shall discuss different types of waves in fluids, depending on the geometry and the external forces acting on them. To begin we shall concentrate mostly on waves in liquids. In particular, we shall deal with incompressible fluids. The different sections in the chapter will include discussion on small amplitude waves, the study of waves in

© 2015 by Taylor & Francis Group, LLC

water under the influence of gravity, linear capillary and gravity waves, and surface waves. The derivation of Klein−Gordon and Boussinesq equations is mentioned briefly. The concept of shallow water waves (the depth of the layer being smaller than the wavelength of the wave) is discussed in detail. We will also discuss nonlinear effects of the otherwise small amplitude waves.The special case of plane waves in a layer of constant depth is dealt with briefly. We shall present some examples as applicable to geophysical fluid dynamics (Poincaré, Kelvin, inertial waves). Application of Rayleigh and Lamb waves to hydrodynamics and soil mechanics is mentioned in passing. Waves which have large wavelengths have to be discussed in a rotating frame. Thus, we shall discuss waves, both in nonrotating as well as rotating systems.

A detailed derivation of the Solitary waves (the evolution equation being the famous Kortweg−de Vries) is made at the end of the chapter. Simple examples of KdV equations and solutions are mentioned briefly. Interested readers are advised to go through the classic works of Whitham [174], Lighthill [142], Shivamoggi [164], Landau and Lifschitz [140], and Pedlosky [152] to learn more about waves in hydrodynamics.

4.2 Basic Equations

The starting points to discuss the dynamical motions are determined by the systematic application of the fluid continuum equations of motion. The dynamic variables which are used to describe the motion are the density ρ, the pressure p, and the velocity vector $\mathbf{u}$. Additionally, certain thermodynamic variables, such as the temperature T, the internal energy per unit mass e, and the specific entropy s, may be included depending on the particular physical problem of interest.

We intend to use the Eulerian kinematic description; unless stated otherwise, the dynamic variables are functions of time t and the vector position coordinate $\mathbf{r}$. In what follows, we shall state the equations of motion in an inertial or nonrotating frame of reference. Subsequently, we shall briefly mention the equations of motion in a rotating frame.

In the absence of sources (s) or sink (s) of mass within the fluid, the conservation of mass is generally expressed by the equation of continuity which is given by

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \rho \mathbf{u} = 0 \tag{4.1}$$

The implication of the above equation is that the local increase in the density with time be balanced by the divergence of the mass flux $\rho \mathbf{u}$. Another way of writing the above equation is

$$\frac{d\rho}{dt} + \rho \boldsymbol{\nabla} \cdot \mathbf{u} = 0 \tag{4.2}$$

© 2015 by Taylor & Francis Group, LLC

where

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u}\cdot\boldsymbol{\nabla}$$

is the total derivative with respect to time of any of the dependent variables related to the individual fluid elements.

The momentum equation is written using Newton's law of motion which is given by

$$\rho\frac{d\mathbf{u}}{dt} = -\boldsymbol{\nabla}p + \rho\boldsymbol{\nabla}\phi + \mathcal{F}(\mathbf{u}) \tag{4.3}$$

Physically, the above equation implies that the mass per unit volume times the acceleration is equal to the sum of the pressure gradient force, the body force $\rho\boldsymbol{\nabla}\phi$, and the $\mathcal{F}$, where ϕ is the potential by which conservative forces such as gravity may be represented. $\mathcal{F}$ in general is a nonconservative force; typically it is related to the frictional force in the fluid. In the case of Newtonian fluids such as air or water,

$$\mathcal{F} = \mu\boldsymbol{\nabla}^2\mathbf{u} + \frac{\mu}{3}\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{u}) \tag{4.4}$$

where μ is the molecular viscosity. $\mathcal{F}$ may be considered to be exact when μ is a function of the thermodynamic state variables, which are constant over the field of motion.

For interesting phenomena wherein the density is considered a function of time, the momentum and continuity equations are insufficient to describe the dynamical systems. Another equation which involves the laws of thermodynamics needs to be incorporated to complete the system. In the first instance, let us look at the first law of thermodynamics which is written as

$$\rho\frac{de}{dt} = -p\rho\frac{d}{dt}\rho^{-1} + \kappa\boldsymbol{\nabla}^2T + \chi + \rho Q \tag{4.5}$$

where e is the internal energy per unit mass, T is the temperature, κ is the thermal conductivity, Q is the rate of heat addition per unit mass by internal heat sources, and χ is the addition of heat due to viscous dissipation. For practical purposes, the change in the internal energy due to viscous dissipation may be neglected.

In addition to the above equations, we may have to introduce the thermodynamic state property s, the specific entropy. The relation between the entropy and other variables is given by

$$T\Delta s = \Delta e = p\Delta(1/\rho) \tag{4.6}$$

where Δs, Δe, and $\Delta(1/\rho)$ are arbitrary increments in s, e, and $1/\rho$. In particular,

$$T\frac{ds}{dt} = \frac{de}{dt} + p\frac{d}{dt}\left(\frac{1}{\rho}\right) \tag{4.7}$$

The equation involving the thermodynamic variables may be rewritten as

$$T\frac{ds}{dt} = \frac{k}{\rho}\boldsymbol{\nabla}^2T + Q \tag{4.8}$$

© 2015 by Taylor & Francis Group, LLC

In order to complete the system of equations, the relation among the density, pressure, and temperature, called the equation of state, needs to be included and may be written as

$$\rho = \rho(p, T) \tag{4.9}$$

and

$$s = s(p, T) \tag{4.10}$$

4.2.1 Equations in a Rotating Frame

For a rotating coordinate frame, the relation between the vector $\mathbf{A}$ in nonrotating and rotating frames is given by

$$\left(\frac{d\mathbf{A}}{dt}\right)_I = \left(\frac{d\mathbf{A}}{dt}\right)_R + \Omega \times \mathbf{A} \tag{4.11}$$

The subscripts I and R denote the changes as seen by the observer in the nonrotating and rotating frames, respectively. Ω is the angular velocity of the rotating observer. The above analogy may be extended to a position vector $\mathbf{r}$ of an arbitrary fluid element. We may simply write the relation as

$$\mathbf{u}_I = \mathbf{u}_R + \Omega \times \mathbf{r} \tag{4.12}$$

The rate of change of velocity is proportional to the applied forces per unit mass to the acceleration in inertial space, which means

$$\left(\frac{d\mathbf{u}_I}{dt}\right)_I = \left(\frac{d\mathbf{u}_I}{dt}\right)_R + \Omega \times \mathbf{u}_I \tag{4.13}$$

The right-hand side of the above equation may be simplified to yield

$$\left(\frac{d\mathbf{u}_I}{dt}\right)_I = \left(\frac{d\mathbf{u}_R}{dt}\right)_R + 2\Omega \times \mathbf{u}_R + \Omega \times (\Omega \times \mathbf{r}) + \frac{d\Omega}{dt} \times \mathbf{r} \tag{4.14}$$

The momentum equation for an observer in a uniformly rotating coordinate frame may be simplified to yield

$$\rho \left[\frac{d\mathbf{u}}{dt} + 2\Omega \times \mathbf{u}\right] = -\boldsymbol{\nabla} p + \rho \boldsymbol{\nabla} \Phi + \mathcal{F} \tag{4.15}$$

4.3 Small-Amplitude Waves

In this section, we shall study the behavior of the dynamical equations for small-amplitude motions. For this we need to linearize the equations of

© 2015 by Taylor & Francis Group, LLC

motion. Let us consider only two-dimensional time-dependent, incompressible fluids with gravity g as the restoring force.

Consider a fluid layer of thickness in the absence of motions to be $H_0(x, y)$. We perturb this expression as

$$H(x, y, t) = H_0(x, y) + \eta(x, y, t) \tag{4.16}$$

Here $\eta(x, y, t)$ is the perturbation in the fluid layer. For small-amplitude motions, one imposes the constraint that $\eta \ll H_0$. In addition, we suppose that the velocities u and v are small such that

$$\frac{\partial \mathbf{u}_H}{\partial t} \gg \mathbf{u}_H \cdot \boldsymbol{\nabla} \mathbf{u}_H \tag{4.17}$$

Ignoring the quadratic terms in the dynamical variables, the linearized system of equations can be written as

$$\frac{\partial u}{\partial t} - fv = -g\frac{\partial \eta}{\partial x} \tag{4.18}$$

$$\frac{\partial v}{\partial t} + fu = -g\frac{\partial \eta}{\partial y} \tag{4.19}$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x}(uH_0) + \frac{\partial}{\partial y}(vH_0) = 0 \tag{4.20}$$

Defining the linearized mass flux vector $\mathbf{U} = iU + jV$, where $U = uH_0$ and $V = vH_0$, the above equations can be simplified to yield

$$\frac{\partial U}{\partial t} - fV = -gH_0\frac{\partial \eta}{\partial x} \tag{4.21}$$

$$\frac{\partial V}{\partial t} + fU = -gH_0\frac{\partial \eta}{\partial y} \tag{4.22}$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} = 0 \tag{4.23}$$

Here, f is a measure of the rotation of the fluid. With simple algebra, equations (4.21 and 4.22) can be simplified to yield

$$\frac{\partial}{\partial t}\left[\frac{\partial V}{\partial y} + \frac{\partial U}{\partial x}\right] - f\left[\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right] = -g\boldsymbol{\nabla}\cdot[H_0\boldsymbol{\nabla}\eta] \tag{4.24}$$

$$\frac{\partial}{\partial t}\left[\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right] + f\left[\frac{\partial V}{\partial y} + \frac{\partial U}{\partial x}\right] = -g\left[\frac{\partial H_0}{\partial x}\frac{\partial \eta}{\partial y} - \frac{\partial H_0}{\partial y}\frac{\partial \eta}{\partial x}\right] \tag{4.25}$$

The equations (4.24 and 4.25) can be simplified to yield

$$\left[\frac{\partial^2}{\partial t^2} + f^2\right]\left[\frac{\partial V}{\partial y} + \frac{\partial U}{\partial x}\right] = -g\frac{\partial}{\partial t}\boldsymbol{\nabla}\cdot(H_0\boldsymbol{\nabla}\eta) - fg\left[\frac{\partial H_0}{\partial x}\frac{\partial \eta}{\partial y} - \frac{\partial H_0}{\partial y}\frac{\partial \eta}{\partial x}\right] \tag{4.26}$$

© 2015 by Taylor & Francis Group, LLC

A single equation for η can be obtained by using equations (4.23) and (4.26) to yield

$$\frac{\partial}{\partial t}\left[\left(\frac{\partial^2}{\partial t^2}+f^2\right)\eta-\boldsymbol{\nabla}\cdot(C_0^2\boldsymbol{\nabla}\eta)\right]-gfJ(H_0,\eta)=0 \tag{4.27}$$

where $C_0^2=gH_0$ and $J(A,B)$ denotes the Jacobian of the two functions. The velocities u and v can be written in terms of η. Once the form for η is derived by solving equation (4.27), the expressions for u and v can be determined.

4.3.1 An Application in Geophysics

In what follows, we give a simple example of the above dynamical equations, with an application to geophysics. Consider the system wherein the equations of motion are assumed to be time-independent to start with. Assuming that there is no time variation for η in equation (4.27), we can rewrite them with the following equations:

$$u = -\frac{g}{f}\frac{\partial\eta}{\partial y} \tag{4.28}$$

$$v = \frac{g}{f}\frac{\partial\eta}{\partial x} \tag{4.29}$$

The above relations, are the geostrophic relations for the horizontal motions. Looking at the similarity with the Euler equations, one can realize that the pressure and η are related to each other. For example, the isolines of η are nothing but the streamlines for the steady geostrophic flow as

$$u\frac{\partial\eta}{\partial x}+v\frac{\partial\eta}{\partial y}=0 \tag{4.30}$$

The steady assumption leads to a much simplified equation for η, namely,

$$J(H_0,\eta)=0 \tag{4.31}$$

where J is the Jacobian of the system. The physical implication is that lines of constant undisturbed depth H_0 must coincide with lines of constant η in the x,y plane. Thus, if η is specified at one point on each H_0, then it can be determined everywhere on that contour.

It should be noted that real motions in general are not precisely geostrophic, and that one cannot have constraints such as steadiness and linearity.

4.4 Gravity Waves

For an incompressible fluid, even in the absence of a free surface, buoyancy forces (those due to the presence of gravity) can give rise to gravity waves

© 2015 by Taylor & Francis Group, LLC

(sometimes called internal gravity waves) if different fluid elements have different concentrations and densities. These waves are generally found in oceans. One can in principle set up laboratory experiments for studying such waves. Let us assume that the fluid has a stratification in the density, which is merely a function of the vertical coordinate, i.e., $\rho_0 = \rho_0(y)$ for a two-dimensional fluid. The hydrostatic pressure balance leads to an expression for the variation in the pressure as

$$0 = -\frac{dp_0}{dy} - \rho_0 \mathbf{g} \tag{4.32}$$

Assume that the fluid is disturbed slightly from this equilibrium state. It is important to realize that each element conserves its mass, its volume, and hence its density. The governing equations are the standard Euler's equation written as

$$\rho\left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{u}\right] = -\boldsymbol{\nabla} p + \rho \mathbf{g} \tag{4.33}$$

$$\boldsymbol{\nabla}\cdot\mathbf{u} = 0 \tag{4.34}$$

$$\frac{\partial \rho}{\partial t} + (\mathbf{u}\cdot\boldsymbol{\nabla})\rho = 0 \tag{4.35}$$

For a two-dimensional flow with small amplitude, one can write

$$\mathbf{u} = [u(x,y,t), v(x,y,t), 0]; \quad p = p_0(y) + p_1(x,y,t); \quad \rho = \rho_0(y) + \rho_1(x,y,t)$$

Substituting the above expressions into the equations of motion (4.33−4.35) and neglecting the quadratically small terms in u_1, v_1, and p_1, we obtain

$$\rho_0 \frac{\partial u_1}{\partial t} = -\frac{\partial p_1}{\partial x} \tag{4.36}$$

$$\rho_0 \frac{\partial v_1}{\partial t} = -\frac{\partial p_1}{\partial y} - \rho_1 g \tag{4.37}$$

$$\frac{\partial u_1}{\partial x} + \frac{\partial v_1}{\partial y} = 0 \tag{4.38}$$

$$\frac{\partial \rho_1}{\partial t} + v_1 \frac{d\rho_0}{dy} = 0 \tag{4.39}$$

Assuming perturbations for the flow variables as

$$v_1 = \hat{v_1}(y) e^{i(kx - \omega t)}$$

the equations of motion (linearized) can be written as

$$\rho_0 \omega \hat{u_1} = k\hat{p_1} \tag{4.40}$$

$$\rho_0 i\omega \hat{v_1} = \hat{p_1}' + \hat{\rho} g \tag{4.41}$$

$$ik\hat{u_1} + \hat{v_1}' = 0 \tag{4.42}$$

$$-i\omega\hat{\rho_1} + \hat{v_1}\frac{d\rho_0}{dy} = 0 \tag{4.43}$$

© 2015 by Taylor & Francis Group, LLC

Here the dash ($'$) denotes differentiation with respect to y. Eliminating the other variables, retaining only v_1, leads to a single differential equation as follows:

$$\hat{v_1}'' + \frac{\rho_0'}{\rho_0}\hat{v_1'} + k^2\left(\frac{N^2}{\omega^2} - 1\right)\hat{v_1} = 0 \tag{4.44}$$

The quantity N is defined as the buoyancy frequency (also called Brunt–Vaisala frequency) in the literature, given by

$$N^2 = -\frac{g}{\rho_0}\frac{d\rho_0}{dy} \tag{4.45}$$

If we also assume that the density decreases exponentially with height, so that $\rho_0 \propto e^{-y/H}$, then the buoyancy frequency N and the coefficients of the differential equation become constant. Thus, we can solve the equation for wave-like solutions for v_1 of the form

$$v_1 \propto e^{y/2H}e^{i(kx-\omega t)}$$

where

$$\omega^2 = \frac{N^2k^2}{(k^2 + l^2 + 1/4H^2)}$$

The propagation of the internal gravity waves is, as is clear from the dispersion relation mentioned above, anisotropic because the frequency ω depends on k and l. For the special case when the wavelength $\lambda = 2\pi/(k^2 + l^2)^{1/2}$ is small as compared to the scale height H, the dispersion relation reduces to

$$\omega^2 = \frac{N^2k^2}{(k^2 + l^2)}$$

The group velocity (see Chapter 1 also) is given by

$$c_g = \left(\frac{\partial\omega}{\partial k}, \frac{\partial\omega}{\partial l}\right)$$

Differentiating the dispersion relation yields the expression for the group velocity as

$$c_g = \frac{\omega l}{(k^2 + l^2)k}(l, -k) \tag{4.46}$$

The lines of constant phase are given by the equation $kx + ly - \omega t = constant$. It is interesting to note that the crests move in the direction of (k, l), while the group velocity moves in a perpendicular direction, $(l, -k)$. This is typical behavior of 2-dimensional gravity waves.

© 2015 by Taylor & Francis Group, LLC

4.4.1 Waves in a Steady Stream

We shall study waves in a steady stream. Assume the applied pressure on $z = 0$ to be

$$\frac{p - p_0}{\rho} = f(x, y)e^{\omega t} \tag{4.47}$$

The equations of motion can be written as

$$\begin{aligned}
\boldsymbol{\nabla}^2\phi &= 0, \quad z < 0 && (4.48)\\
z &= 0, \quad \psi_t + U\psi_x = \phi_x && (4.49)\\
z &= 0, \quad \phi_t + U\phi_x + g\psi - \frac{T}{\rho}(\psi_{xx} + \psi_{yy}) = -f(x, y)e^{\omega t} && (4.50)\\
z &\to -\infty, \quad \phi \to 0 && (4.51)
\end{aligned}$$

ψ in the above equations is the displacement. Introducing the following transformations:

$$\phi = e^{\omega t}\int_{-\infty}^{\infty} B(\mathbf{K})e^{i(k_x x + k_y y) + kz}d\mathbf{K}$$

$$\psi = e^{\omega t}\int_{-\infty}^{\infty} A(\mathbf{K})e^{i(k_x x + k_y y)}d\mathbf{K}$$

$$f = e^{\omega t}\int_{-\infty}^{\infty} F(\mathbf{K})e^{i(k_x x + k_y y)}d\mathbf{K}$$

where $k = \sqrt{k_x^2 + k_y^2}$. The solution for ψ can be found using Fourier transform as

$$\psi(x, t) = e^{\omega t}\int_{-\infty}^{\infty} \frac{kF(\mathbf{K})e^{i(k_x x + k_y y)}d\mathbf{K}}{(k_x U - i\omega)^2 - \omega_0^2(k)} \tag{4.52}$$

where $\omega_0^2 = gk + \frac{T}{\rho}k^3$.

4.5 Linear Capillary and Gravity Waves

In this section we shall consider two-dimensional surface waves in a frictionless, homogeneous, and incompressible fluid, with a flat and horizontal bottom as shown in Figure 4.1. The x axis forms the surface while the fluid is in and rest. Assume the depth of the fluid to be H. The z axis is the vertical axis.

The wave motion is assumed to be nonrotational so that the velocity potential ϕ satisfies the Laplace equation given by

$$\boldsymbol{\nabla}^2\phi = 0 \tag{4.53}$$

© 2015 by Taylor & Francis Group, LLC

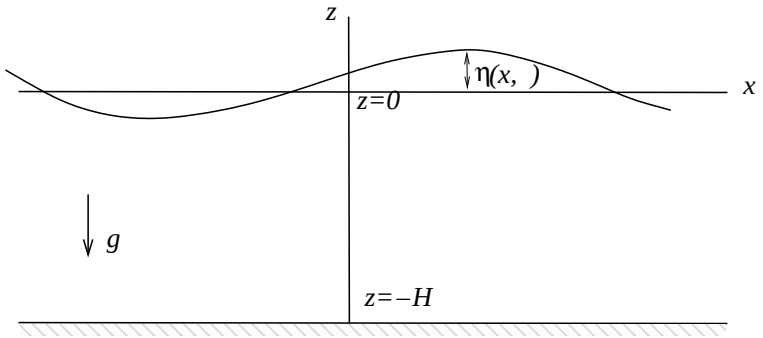


FIGURE 4.1: Sketch for a two-dimensional wave.

The boundary conditions (in the linear limit) are the following:

$$\frac{\partial \phi}{\partial t} + g\eta - \frac{\sigma}{\eta}\frac{\partial^2 \eta}{\partial x^2} = 0 \tag{4.54}$$

$$\frac{\partial \phi}{\partial z} = \frac{\eta}{\partial t} \tag{4.55}$$

for $z = 0$. The quantity σ is a measure of the surface tension of the free surface. The additional boundary condition at the bottom, $z = -H$,

$$\frac{\partial \phi}{\partial z} = 0 \tag{4.56}$$

The solution of the Laplace equation, subject to the boundary conditions mentioned above can be calculated and is given by

$$\eta = A\sin(kx - \omega t) \tag{4.57}$$

$$\phi = \frac{A\omega}{k\sinh kH}\cosh k(z+H)\cos(kx - \omega t) \tag{4.58}$$

where A is the amplitude of the wave. The wavenumber and angular frequency ω are related to each other by the dispersion relation given by

$$\omega^2 = \left(gk + \frac{\sigma k^3}{\rho}\right)\tanh kH \tag{4.59}$$

The phase speed for the capillary and gravity wave is given by

$$C = C_0\left(1 + \frac{\sigma k^2}{\rho g}\right)^{1/2}\left(\frac{\tanh kH}{kH}\right)^{1/2} \tag{4.60}$$

where $C_0 = \sqrt{gH}$. The effect of surface tension is negligible in the above

© 2015 by Taylor & Francis Group, LLC

expression for the case when $\sigma k^2/\rho g \ll 1$. Thus, the surface tension may be neglected for waves whose wavelengths are much larger than λ_m, where

$$\lambda_m = 2\pi \left(\frac{\sigma}{\rho g}\right)^{1/2} \tag{4.61}$$

On the other hand, the effect of surface tension will be more important and will dominate the wave motion when the wavelength is much smaller than λ_m. For the case when the wavelength is much less than H, i.e., $kH \gg 1$, $\tanh kH = 1$, which means

$$C = \left(\frac{g}{k} + \frac{\sigma k}{\rho}\right)^{1/2} \tag{4.62}$$

The phase velocity has a minimum value given by

$$C_m = 2^{1/2} \left(\frac{\sigma g}{\rho}\right) \tag{4.63}$$

The ratio of C/C_m has the form

$$\frac{C}{C_m} = \left[\frac{1}{2}\left(\frac{\lambda}{\lambda_m} + \frac{\lambda_m}{\lambda}\right)\right]^{1/2} \tag{4.64}$$

It is interesting to note that when the wavelength is larger than λ_m, the propagation velocity is larger for long waves (waves whose wavelength is large compared to the depth of the fluid), while the opposite trend is exhibited for wavelengths shorter than λ_m. The dispersive characteristics are shown in Figure 4.2, where C is shown as a function of the wavelength.

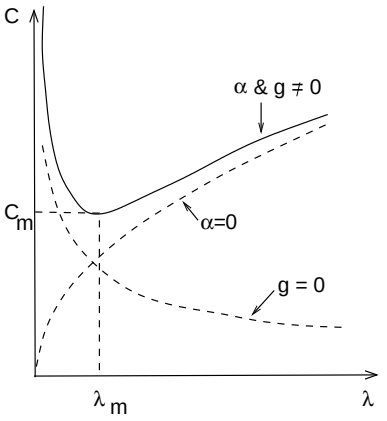


FIGURE 4.2: Dispersive characteristics: C as a function of the wavelength λ.

© 2015 by Taylor & Francis Group, LLC

One can expand $\tanh kH$ and approximate it as

$$\tanh kH = kH - \frac{(kH)^3}{3} \tag{4.65}$$

Finally, with simple manipulation we get

$$C = C_0 \left[1 + \left(\frac{\sigma}{\rho g H^2} - \frac{1}{3}\right)\frac{k^2 H^2}{2}\right] \tag{4.66}$$

The linear theory is perfectly valid as long as $A/\lambda \ll 1$ when $kH \gg 1$ and $A/H \ll 1$ when $kH \ll 1$. This means that the steepness of the wave is adequately small.

4.5.1 One-Dimensional Capillary – Gravity Waves

We shall briefly mention an example for one-dimensional capillary-gravity waves. In the discussion on waves in a steady stream, we defined a function f. We will introduce a new variable as

$$f(x) = P\delta(x), \quad F(k_x) = P/2\Pi$$

where $\delta(x)$ is the Dirac delta function.

The solution in this case can be written as

$$\psi(x,t) = \frac{Pe^{\omega t}}{2\Pi}\int_{-\infty}^{\infty} \frac{k_x e^{ik_x x}}{(k_x U - i\omega)^2 - gk_x} dk_x$$

For capillary gravity waves, surface tension plays a very important role. Including the surface tension T and simplifying yields

$$\psi(x,t) = \frac{P}{2\Pi}\lim_{\epsilon\to 0}\int_{-\infty}^{\infty} \frac{e^{ik_x x}dk_x}{k_x U^2 - 2i\omega U - g - \frac{T}{\rho}k_x^2}$$

The poles of the above integrals lie at

$$k_x U^2 \approx g + \frac{T}{\rho}k_x^2$$

Writing in terms of k_x, the poles will lie at two points:

$$k_x = k_{xg} + \frac{2\rho U}{(k_{xT} - k_{xg})T}i\omega, \quad k_x = k_{xT} - \frac{2\rho U}{(k_{xT} - k_{xg})T}i\omega$$

k_{xg} and k_{xT} are the two real roots with the condition that $k_{xT} > k_{xg}$.

© 2015 by Taylor & Francis Group, LLC

4.6 Surface Waves Generated by a Local Disturbance in the Field

Whenever a stone is thrown into still and relatively deep water, one sees a train of regular waves, which tend to propagate radially outwards from the point where the stone was thrown. If, for example, the stone was large, we can notice long waves in the front of the wave train, followed by shorter waves towards the center of the motion. The physical explanation is that the mechanical energy which contributes to the motion in the water is distributed to the different wave components, which propagate with a velocity which depends on the wavelength. With the passage of time, the long waves will overtake the shorter ones, with the result that the original disturbance is transformed into a regular wave train with gradually varying wavelengths. However, if a stone is tossed into a shallow water, then all the wave components whose wavelengths are greater than the depth of the water will move with approximately the same velocity. The disturbance does not suffer in its form over long distances. When the stone is small, the long waves generated will have very small amplitudes such that only those waves whose wavelengths $\lambda \leq \lambda_m$ will be visible.

We shall try to model such a scenario. For the sake of simplicity, we shall restrict ourselves to two-dimensional wave motion, which has arisen from elongated disturbance. This is relatively easier to model mathematically compared to the waves which propagate in all the directions such that the amplitude spreads out over a large area. The starting point is the equation (4.53):

$$\boldsymbol{\nabla}^2 \phi = 0 \tag{4.67}$$

with the following conditions:

$$\phi = 0 \quad \text{for} \quad t = 0 \tag{4.68}$$

$$\eta = \eta_0(x) \quad \text{for} \quad t = 0 \tag{4.69}$$

Define the Fourier transform with respect to x as

$$\hat{f}(k) = \int_{-\infty}^{+\infty} f(x) e^{-ikx} dx \tag{4.70}$$

where $f(x)$ is a function of x such that the above integral exists. The inverse of the above transform is given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(k) e^{ikx} dk \tag{4.71}$$

We also assume that ϕ and the derivative η tends to zero for $x \to \pm\infty$ and

© 2015 by Taylor & Francis Group, LLC

that the Fourier transform of these functions exists. Integrating by parts one realizes that

$$\widetilde{\frac{\partial^n \phi}{\partial x^n}} = (ik)^n \hat{\phi} \quad \text{for} \quad (n = 1, 2, \cdots) \tag{4.72}$$

The Fourier transform for the wave equation can be written as

$$\frac{\partial^2 \hat{\phi}}{\partial z^2} - k^2 \hat{\phi} = 0 \tag{4.73}$$

The transform for the first of the boundary conditions is given by

$$\hat{\eta} = -\frac{1}{g\left(1 + \frac{\sigma k^2}{\rho g}\right)} \left(\frac{\partial \hat{\phi}}{\partial t}\right) = 0 \tag{4.74}$$

Eliminating $\hat{\eta}$ with one of the boundary conditions yields

$$\frac{\partial^2 \hat{\phi}}{\partial t^2} + g\left(1 + \frac{\sigma k^2}{\rho g}\right) \frac{\partial \hat{\phi}}{\partial z} = 0 \quad \text{for} \quad z = 0 \tag{4.75}$$

The transform for the boundary condition is given by

$$\frac{\partial \hat{\phi}}{\partial z} \quad \text{for} \quad z = -H \tag{4.76}$$

Finally, the initial conditions have the following Fourier transforms:

$$\hat{\phi} = 0 \quad \text{and} \quad \hat{\eta} = \hat{\eta_0} \quad \text{for} \quad t = 0 \tag{4.77}$$

We seek a solution for the equation in the form

$$\hat{\phi} = A(t) \cosh k(z + H) \tag{4.78}$$

where $A(t)$ is an undefined function of t. The differential equation for $\hat{\phi}$ with the above assumption reduces to

$$\frac{d^2 A}{dt^2} = -\omega^2 A \tag{4.79}$$

where ω is given by equation (4.59).

Applying the initial conditions, one gets $A = 0$ for $t = 0$, so that the solution of equation (4.79) may be written as

$$A(t) = A_0 \sin \omega t \tag{4.80}$$

where the constant A_0 is determined as

$$A_0 = -\frac{g\left(1 + \frac{\sigma k^2}{\rho g}\right)}{\omega \cosh kH} \hat{\eta_0} \tag{4.81}$$

© 2015 by Taylor & Francis Group, LLC

The velocity potential can be simplified to yield

$$\phi = -\frac{g}{2\pi}\int_{-\infty}^{+\infty}\frac{\sin\omega t}{\omega}\frac{\cosh k(z+H)}{\cosh kH}\left(1+\frac{\sigma k^2}{\rho g}\right)\hat{\eta_0}e^{ikx}dk \tag{4.82}$$

Setting the velocity potential and the corresponding boundary condition for ϕ, after integration the surface displacement may be written as

$$\eta = \frac{1}{4\pi}\left[\int_{-\infty}^{+\infty}\hat{\eta_0}e^{i(kx+\omega t)}dk + \int_{-\infty}^{+\infty}\hat{\eta_0}e^{i(kx-\omega t)}dk\right] \tag{4.83}$$

The complex conjugate of $\hat{\eta_0}(k)$ is $\hat{\eta_0}(k)$ and $\omega(-k) = \omega(k)$. Thus, the integrals can be converted in such a way that one integrates over positive values of k. Then we have

$$\eta = \frac{1}{2\pi}\int_0^{\infty}|\hat{\eta_0}(k)|(\cos(kx-\omega t+\gamma)+\cos(kx+\omega t+\gamma))dk \tag{4.84}$$

where γ is the argument of $\hat{\eta_0}$.

The above expression implies that the motion is made up of a spectrum of harmonic wave components. The original disturbance is split up into wave components which have both positive and negative components in the x direction. The complex Fourier transform $\hat{\eta_0}$ will determine the phase and amplitude of the different wave components. The wave spectrum may be defined as the square of the modulus $|\hat{\eta_0}(k)|^2$. The Fourier integral can be inverted to yield

$$\eta = \frac{1}{2}\eta_0(x+C_0t) + \frac{1}{2}\eta_0(x-C_0t) \tag{4.85}$$

The above expression implies that the original disturbance splits into two identical pulses which move with constant velocity in both the positive and negative x directions without changing the form.

Define a new transformation given by

$$\eta_0(x) = \frac{Q}{2L\sqrt{\pi}}e^{-(x/2L)^{1/2}} \tag{4.86}$$

where Q and L are constants.

The important properties are as follows:

$$\int_{-\infty}^{+\infty}\eta_0(x)dx = Q \tag{4.87}$$

$$\hat{\eta_0}(x) = Qe^{-(kL)^2} \tag{4.88}$$

The modulus $|\hat{\eta_0}|$ is symmetric around $k=0$. For $L\to 0$, one can show that $\hat{\eta_0}(x)$ tends to $Q\delta(x)$, where $\delta(x)$ is the famous Dirac delta function.

Let us use the properties of the delta function which is

$$\int_{-\infty}^{+\infty}\delta(x)dx = 1 \quad \delta(x) = 0 \quad \text{for} \quad x \neq 0 \tag{4.89}$$

© 2015 by Taylor & Francis Group, LLC

and

$$\hat{\delta}(k) = 1$$

An important property of the delta function is that it has a spectrum which gives equal amplitude for all the wave components. As already mentioned above, the surface displacement which is symmetric around the origin can be written as

$$\eta = \frac{Q}{2\pi} \int_0^{+\infty} e^{-(kL)^2} [\cos(kx + \omega t) + \cos(kx - \omega t)] dk \tag{4.90}$$

Let us introduce a new variable u such that

$$\omega = (\frac{g}{x})^{1/2}(u \pm r)$$

$$k = \frac{1}{x}(u^2 \pm 2ur + r^2)$$

where

$$kx \pm \omega t = u^2 - r^2$$

and

$$r = \left(\frac{gt^2}{4x}\right)^{1/2}$$

In the limit $L \to 0$:

$$\eta = \frac{2Q}{\pi} \frac{r}{x} \int_0^r \cos(u^2 - r^2) du \tag{4.91}$$

The surface displacement can be determined with the help of the boundary conditions. For more details, the reader may refer to the book by Lamb [139].

4.7 Klein−Gordon Equation

Klein−Gordon equation is an interesting equation which describes the evolution in addition to waves in hydrodynamics (for example, long gravity waves in rotating fluids), the oscillating string, relativistic quantum mechanics. We shall skip the derivation of this equation in this section for brevity. However, we will discuss more about its implications in magneto-hydrodynamic waves in the next chapter. To begin with, we shall look at the form of the equation and some salient features of its solution. The differential equation may be written as

$$\frac{\partial^2 \eta}{\partial t^2} - C_0^2 \frac{\partial^2 \eta}{\partial x^2} + C_0^2 q^2 \eta = 0 \tag{4.92}$$

© 2015 by Taylor & Francis Group, LLC

Here, we assume that C_0 and q are constant, while η is a function of x and t. We seek simple solutions of the form

$$\eta = A \sin k(x - Ct)$$

where the phase velocity is

$$C = C_0(1 + q^2/k^2)^{1/2}$$

The group velocity is given by

$$C_g = \frac{d(Ck)}{dk} = \frac{C_0^2}{C}$$

Multiplying the Klein−Gordon equation by $\partial\eta/\partial t$ and reordering, we obtain

$$\frac{\partial E}{\partial t} + \frac{\partial F}{\partial x} = 0 \tag{4.93}$$

where the energy density E is given by

$$E = \frac{1}{2}\left[\left(\frac{\partial \eta}{\partial t}\right)^2 + C_0^2\left(\frac{\partial \eta}{\partial x}\right)^2 + C_0^2 q^2 \eta^2\right]$$

and the energy flux is

$$F = -C_0^2 \frac{\partial \eta}{\partial x}\frac{\partial \eta}{\partial t}$$

The average propagation velocity for the energy is given by

$$\frac{\bar{F}}{\bar{E}} = \frac{C_0^2}{C} = C_g \tag{4.94}$$

4.8 Shallow Water Waves

We assume that the fluid is incompressible so that the density decouples from the dynamical equations of thermodynamics. The mass conservation equation can be written as

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{4.95}$$

For shallow water equations, we also assume that the flow variables u and v are independent of z. The departure from the hydrostatic approximation may be modeled as

$$p = -\rho g z + A(x, y, t) \tag{4.96}$$

© 2015 by Taylor & Francis Group, LLC

Introducing the boundary condition $p(x, y, h) = p_0$, where p_0 is a constant, leads to

$$p = \rho g(h - z) + p_0 \tag{4.97}$$

The above relation implies that the pressure in excess of p_0 at any point is equal to the weight of the unit column of fluid above the point at that instant. With the assumption that the horizontal pressure gradient is independent of z, we see that the horizontal accelerations are also independent of z. The equations for horizontal pressure gradient are given by

$$\frac{\partial p}{\partial x} = \rho g \frac{\partial h}{\partial x} \tag{4.98}$$

$$\frac{\partial p}{\partial y} = \rho g \frac{\partial h}{\partial y} \tag{4.99}$$

The above equations are consistent with the fact that the horizontal velocities are also independent of z. The famous Taylor−Proudman theorem asserts that for a homogeneous fluid which rotates uniformly, the flow velocities are independent of z. Thus, the horizontal momentum equations can be written as

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv = -g\frac{\partial h}{\partial x} \tag{4.100}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu = -g\frac{\partial h}{\partial y} \tag{4.101}$$

Integrating the equation of continuity with the assumption that u and v are independent of z leads to

$$w(x, y, z, t) = -z\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \hat{w}(x, y, t) \tag{4.102}$$

The boundary condition that the normal flow at a rigid surface $z = h_B$ requires that

$$w(x, y, h_B, t) = u\frac{\partial h_B}{\partial x} + v\frac{\partial h_B}{\partial y} \tag{4.103}$$

Therefore,

$$\hat{w}(x, y, t) = u\frac{\partial h_B}{\partial x} + v\frac{\partial h_B}{\partial y} + h_B\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{4.104}$$

so that

$$w(x, y, z, t) = (h_B - z)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + u\frac{\partial h_B}{\partial x} + v\frac{\partial h_B}{\partial y} \tag{4.105}$$

The corresponding kinematic condition at the surface is given by $z = h$:

$$w = \frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y}, \qquad z = h(x.y, t) \tag{4.106}$$

© 2015 by Taylor & Francis Group, LLC

Eliminating w between equation (4.105) and (4.106) yields

$$\frac{\partial h}{\partial t} + \frac{\partial h}{\partial x}[(h - h_B)u] + \frac{\partial}{\partial y}[(h - h_B)v] = 0 \tag{4.107}$$

Denoting the total depth by $H = h - h_B$, the mass conservation equation becomes

$$\frac{\partial H}{\partial t} + \frac{\partial}{\partial x}(uH) + \frac{\partial}{\partial y}(vH) = 0 \tag{4.108}$$

The set of equations defined by (4.100, 4.101, and 4.108) is the basis for shallow water equations. It is assumed that $\delta = D/L$, where D and L are vertical and horizontal length scales is very much less than unity (see Figure 4.3).

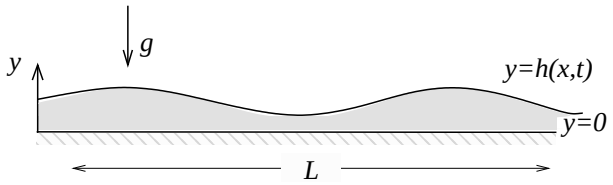


FIGURE 4.3: Simple sketch for a shallow water model.

4.8.1 Long Waves in Shallow Water

For a wave whose wavelength is large in comparison with the depth and whose amplitude is sufficiently small, the water particles move in such a way that the horizontal velocity is much longer than the vertical velocity. Such waves are termed long waves in shallow water. The horizontal velocity is approximately constant as is the vertical shape, whereas the vertical acceleration is comparable to the acceleration due to gravity. For such a scenario, one can assume that the pressure of the fluid is approximately in hydrostatic equilibrium, with a value close to the pressure overlying the fluid. With these assumptions, we shall derive the equation for the propagation of long waves in shallow water. Also, we assume that the wavelength is much larger than half of the depth.

Define the horizontal volume flux as

$$U = \int_{-H}^{\eta} u dz \qquad V = \int_{-H}^{\eta} v dz \tag{4.109}$$

For a fluid which is incompressible, the net horizontal disruption in the fluid column must equal the displacement of the surface. This can be denoted by

$$\frac{\partial \eta}{\partial t} = -\frac{\partial U}{\partial x} - \frac{\partial V}{\partial y} \tag{4.110}$$

The above equation can be derived by integrating the equation of continuity

© 2015 by Taylor & Francis Group, LLC

$\nabla \cdot u = 0$ from $z = -H$ to $z = \eta$ and using the kinematic boundary conditions at the bottom and the surface. The above equation still holds if the bottom is not a plane and H is a function of x and y. Since we have assumed that the vertical acceleration is small, the hydrostatic equilibrium, is given by

$$\frac{\partial p}{\partial z} = -\rho g$$

Integrating the above equation yields

$$p = \rho g(\eta - z) + p_0$$

where p_0 is the pressure over the fluid. The basic equations of motion may be written as

$$\frac{\partial u}{\partial t} = -g\frac{\partial \eta}{\partial x} \tag{4.111}$$

$$\frac{\partial v}{\partial t} = -g\frac{\partial \eta}{\partial y} \tag{4.112}$$

Let us introduce the change of variables:

$$u = U/H \quad \text{and} \quad v = V/H$$

This will entail the equations to be written as

$$\frac{\partial U}{\partial t} = -C_0^2\frac{\partial \eta}{\partial x} \tag{4.113}$$

$$\frac{\partial V}{\partial t} = -C_0^2\frac{\partial \eta}{\partial y} \tag{4.114}$$

where $C_0^2 = gH$. The above equations are the linearized shallow water equations. Substituting the relations for $\partial U/\partial t$ and $\partial V/\partial t$ leads to

$$\frac{\partial^2 \eta}{\partial t^2} = \frac{\partial}{\partial x}\left(C_0^2\frac{\partial \eta}{\partial x}\right) + \frac{\partial}{\partial y}\left(C_0^2\frac{\partial \eta}{\partial y}\right) \tag{4.115}$$

If we assume that the bottom boundary is a plane, i.e., C_0, is independent of x and y, then the above equation reduces to the simple two-dimensional wave equation for nondispersive waves. A wave solution can be written as

$$\eta = A\sin k(x - C_0 t)$$

$$u = \frac{A}{H}C_0 \sin k(x - C_0 t)$$

$$v = 0$$

The vertical velocity takes the form

$$w = -\frac{kA}{H}C_0(z + H)\cos k(x - C_0 t)$$

© 2015 by Taylor & Francis Group, LLC

If we assume that the horizontal velocity is independent of z, by integrating the continuity equation, we get

$$w = -\frac{\partial u}{\partial x}(z + H)$$

From the linearized equation of the vertical motion, we find that

$$\frac{\partial p}{\partial z} = -\rho g + \rho \frac{\partial^2 u}{\partial t \partial x}(z + H)$$

Retaining only the linear term, we have

$$p = \rho g(\eta - z) + \rho \frac{\partial^2 u}{\partial t \partial x}\left(\frac{1}{2}z^2 + Hz\right) + p_0 \tag{4.116}$$

The equation for the horizontal motion is

$$\frac{\partial u}{\partial t} = -g\frac{\partial \eta}{\partial x} - \frac{\partial^3 u}{\partial t \partial x^2}\left(\frac{1}{2}z^2 + Hz\right) \tag{4.117}$$

Integrating the above equation in the interval from $z = -H$ to $z = 0$ results in

$$\frac{\partial U}{\partial t} = -C_0^2 \frac{\partial \eta}{\partial x} + \frac{H^2}{3}\frac{\partial^3 U}{\partial t \partial x^2} \tag{4.118}$$

The second term on the right-hand side of the above equation is a correction to the hydrostatic distribution of pressure. We seek solution of the form

$$\eta = A \sin k(x - Ct)$$

$$U = A \sin k(x - Ct)$$

where the phase velocity is

$$C = \frac{C_0}{\left[1 + \frac{(kH)^2}{3}\right]^{1/2}} \approx C_0\left[1 - \frac{(kH)^2}{6}\right] \tag{4.119}$$

The above expression shows that the correction term leads to the dispersion of waves. If the wavelength is sufficiently larger in comparison with H, then we must use the nonlinear term in the equation. The nonlinear plane wave equations in shallow water are given by

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = -g\frac{\partial \eta}{\partial x} \tag{4.120}$$

$$\frac{\partial \eta}{\partial t} = -\frac{\partial}{\partial x}[u(H + \eta)] \tag{4.121}$$

We will have more nonlinear aspects to discuss towards the end of the chapter.

© 2015 by Taylor & Francis Group, LLC

4.9 Boussinesq Equation

In this case, we assume two-dimensional motion with the x axis horizontal and z axis vertical. The bottom boundary need not be a constant and may be described by $z^* = -h^*(x^*)$, where $*$ denotes magnitudes with dimension. We shall use the arguments of nondimensional parameters with the following scales: H is the typical scale for the vertical magnitude, for which z is the vertical velocity and surface displacement. l is the scale used for horizontal magnitudes. The field dimension scales in addition with α which has a dimension of amplitude. Linear theory is restricted to when α is very small. Define a nondimensional parameter as

$$\epsilon = \frac{H^2}{l^2} \tag{4.122}$$

w is assumed to be small. The following nondimensional parameters will be used:

$$z^* = Hz, \quad x^* = lx, \quad t^* = l(gH)^{-1/2}t, \quad h^* = Hh(x).$$
$$\eta^* = \alpha H\eta, \quad u^* = \alpha(gH)^{1/2}u, \quad w^* = \epsilon^{1/2}(gH)^{1/2}, \quad p^* = \rho gHp$$

The boundary conditions are to be imposed at $z = \alpha\eta$, which can be written as

$$p = 0, \quad \eta_t + \alpha u\eta_x \tag{4.123}$$

The subscripts denote differentiation with respect to the variables. At the bottom, the boundary condition will be

$$w = -h_x u \tag{4.124}$$

Euler's equation of motion can be written as

$$u_t + \alpha u u_x + \alpha w u_x = -\alpha^{-1}p_x \tag{4.125}$$
$$\epsilon(w_t + \alpha u w_x + \alpha w w_z) = -\alpha^{-1}(p_x + 1) \tag{4.126}$$

and the equation of continuity is

$$u_x + w_z = 0 \tag{4.127}$$

Comparing the terms of the order ϵ^0 yields

$$p = \alpha\eta - z + O(\alpha\epsilon) \tag{4.128}$$

which means that $p_x = \alpha\eta_x + O(\alpha\epsilon)$ is independent of z to the order of ϵ^0. This implies that all particles in a section $x = const$ have the same horizontal acceleration. They have the same horizontal velocity to the order ϵ^0. One can prove that

$$u_z = O(\epsilon) \tag{4.129}$$

© 2015 by Taylor & Francis Group, LLC

In other words, $u \approx u(x,t)$. The first momentum equation in Euler's equation is written as

$$u_t + \alpha u u_x = -\eta_x + O(\epsilon) \tag{4.130}$$

For the Boussinesq equation, one has to derive a higher-order equation assuming that both ϵ and α are small. This implies that we retain terms of the order of ϵ and α, while products of these or higher-orders can be dropped. In order to proceed further, we define a vertical average as follows:

$$\bar{u} = (h + \alpha\eta)^{-1} \int_{-h}^{\alpha\eta} u dz \tag{4.131}$$

A consequence of the relation (4.130) is that $\hat{u}(x,t) - u(x,z,t) = O(\epsilon)$. An average depth continuity equation can now be written as

$$\eta_t = -[(h + \alpha\eta)\hat{u}]_x \tag{4.132}$$

From the boundary condition for w and the relation for u_x, we have

$$w = \eta_t - z\hat{u}_x + O(\alpha, \epsilon) \tag{4.133}$$

Setting this into the second momentum equation results in

$$p = \alpha\eta - z - \epsilon\alpha\Big(z\eta_{tt} - \frac{1}{2}z^2\hat{u}_{xt}\Big) + O(\alpha\epsilon^2, \alpha^2\epsilon) \tag{4.134}$$

Let us introduce an averaging process as defined below:

$$\overline{(u_t)} = (h + \alpha\eta)^{-1} \int_{-h}^{\alpha\eta} u_t dz = \bar{u}_t + \alpha(h + \alpha\eta)^{-1}(\bar{u} - u|_{z=\alpha\eta})\eta_t \tag{4.135}$$

For the second term in the momentum equation, we have

$$\overline{\Big(\frac{1}{2}\alpha(u^2)_x\Big)} = \alpha\frac{1}{2}(\overline{(u^2)})_x + O(\alpha^2\epsilon) \tag{4.136}$$

Writing $u = \bar{u} + \epsilon u_1$ where u_1 is of the order 1, with the constraint that the average is zero, leads to

$$\frac{1}{2}(\overline{(u^2)})_x = \frac{1}{2}\overline{(\bar{u}^2 + 2\epsilon\bar{u}u_1)_x} + O(\epsilon)^2 \tag{4.137}$$

Combining equations (4.126) and (4.128), we get

$$\bar{u_t} + \alpha\bar{u}\bar{u_x} = -\eta_x + \epsilon\left[\frac{1}{2}h(h\bar{u_t})_{xx} - \frac{1}{6}h^2\bar{u}_{xxt}\right] + O(\epsilon^2, \alpha\epsilon) \tag{4.138}$$

The above equation together with equation (4.126) constitutes a set of two equations in x and t, for the unknowns η and $\bar{u}$. These equations in the literature are termed the Boussinesq equations.

© 2015 by Taylor & Francis Group, LLC

4.10 Finite Amplitude Shallow Water Waves (Nonlinear Aspects)

In what follows we discuss the nonlinear aspects of shallow water waves, which will deal with finite amplitudes, instead of small amplitudes. The long waves (linear theory) may be considered a special case of this section. We consider a two-dimensional study of surface waves, assuming the bottom boundary to be $y = 0$ and a free surface $y = h(x,t)$. For nonlinear studies, it is not correct to assume that the amplitude is small as compared to the depth. Thus, one cannot linearize the equations of motion. Let us assume that h_0 is some typical length scale of $h(x,t)$ and L is a typical horizontal scale length of the wave, such that $h_0 \ll L$. We have already mentioned earlier that this is the condition for shallow water approximation.

Consider the 2-dimensional equations:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial x} \tag{4.139}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - g \tag{4.140}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{4.141}$$

For the shallow water approximation, we can neglect the component of the acceleration Dv/Dt in comparison to g in equations (4.100 and 4.101). The vertical distribution of pressure will satisfy the hydrostatic equilibrium, namely,

$$\frac{\partial p}{\partial y} = -\rho g \tag{4.142}$$

Integrating the above equation and applying the boundary condition, $p = p_0$ at $y = h(x,t)$, we get

$$p = p_0 - \rho g[y - h(x,t)] \tag{4.143}$$

Substituting the above expression into the momentum equation for u, we get

$$\frac{Du}{Dt} = -g\frac{\partial h}{\partial x} \tag{4.144}$$

The above equation implies that the rate of change of u for all fluid elements is independent of y. u being independent of y initially will imply that it will be independent of y all time t. The above equation may be simplified to yield

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = -g\frac{\partial h}{\partial x} \tag{4.145}$$

where u and h are functions of x and t only.

© 2015 by Taylor & Francis Group, LLC

Integrating the equation of continuity (equation 4.95) with respect to y, we get the following expression:

$$v = -\frac{\partial u}{\partial x} y + f(x,t) \tag{4.146}$$

Imposing the condition $v = 0$ at $y = 0$, yields the form for $f(x,t)$ which is zero. The kinematic boundary condition at the free surface is given by

$$v = \frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} \quad at \quad y = h(x,t) \tag{4.147}$$

Combining equations (4.146) and (4.147), we get the relation

$$\frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} + h\frac{\partial u}{\partial x} = 0 \tag{4.148}$$

Equations (4.100) and (4.107) are a couple set of partial differential equations in u and h and they are the starting point for the shallow water equations. Before we discuss further, let us give a justification for neglecting the term Dv/Dt from equation (4.101). Comparing the second and third terms in equation (4.100), we can see that the typical values of u are of the order $(gh_0)^{1/2}$, while with the first term, we realize that $L/(gh_0)^{1/2}$ is a typical time scale. Also we may realize that all terms in equation (4.107) are of the same order. However, considering equation (4.95) which is the equation of continuity, it is easy to realize that v is small compared to u, of the order of $(gh_0)^{1/2}h_0/L$. Also in the momentum equation for v, we find that all the terms are of the order gh_0^2/L^2. These terms are small as compared to g as in the case of shallow water assumption.

We introduce the transformation

$$c(x,t) = (gh)^{/2} \tag{4.149}$$

in place of $h(x,t)$. Equations (4.100) and (4. 107) reduce to

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + 2c\frac{\partial c}{\partial x} = 0 \tag{4.150}$$

$$\frac{\partial (2c)}{\partial t} + u\frac{\partial (2c)}{\partial x} + c\frac{\partial u}{\partial x} = 0 \tag{4.151}$$

With simple algebra the above equations can be simplified to yield

$$\left[\frac{\partial}{\partial t} + (u+c)\frac{\partial}{\partial x}\right](u+2c) = 0 \tag{4.152}$$

$$\left[\frac{\partial}{\partial t} + (u-c)\frac{\partial}{\partial x}\right](u-2c) = 0 \tag{4.153}$$

The above equations are a set of partial differential equations whose solutions can be determined by the method of characteristics.

© 2015 by Taylor & Francis Group, LLC

Define a curve $x = x(s), t = t(s)$ in the $x - t$ plane. Let (x_0, t_0) be some starting point such that $dt/ds = 1$ and $dx/ds = u + c$. With the above choice, equation (4.152) can be written as

$$\left(\frac{dt}{ds}\frac{\partial}{\partial t} + \frac{dx}{ds}\frac{\partial}{\partial x}\right)(u + 2c) = 0 \tag{4.154}$$

Using the chain rule, the above equation reduces to

$$\frac{d}{ds}(u + 2c) = 0 \tag{4.155}$$

The above expression has the simple interpretation that $(u + 2c)$ is a constant along the characteristic curves defined above. With a similar argument for equation (4.153), we can make a statement that $u \pm 2c$ is constant along positive/negative characteristic curves defined by $dx/dt = u \pm c$. This does not completely solve the problem. The final solution would depend on how u and c vary with x and t.

4.11 Plane Waves in a Layer of Constant Depth

In this section, we shall discuss free oscillations in a layer of uniform depth, in such a way that the lateral extent can be idealized as an infinite plane. The extent is assumed to be greater than the wavelength of the oscillations. The layer thickness is assumed to be H_0, similar to the discussion of small amplitude waves section 4.3. In that section, H_0 was assumed to be a function of x and y. In what follows, we shall assume H_0 to be a constant, independent of x and y. The starting point for further discussions is the differential equation (4.27):

$$\frac{\partial}{\partial t}\left[\left(\frac{\partial^2}{\partial t^2} + f^2\right)\eta - \boldsymbol{\nabla}\cdot(C_0^2\boldsymbol{\nabla}\eta)\right] - gfJ(H_0, \eta) = 0$$

With the assumption that H_0 is a constant, the coefficients of the above equation reduce to a constant. We can seek solution to the equation in the form of a plane wave, that is,

$$\eta = Re[\eta_0 e^{i(kx+ly-\omega t)}] \tag{4.156}$$

The symbol Re denotes the real part of the function under consideration. η_0 is the amplitude of the oscillations, while the phase is given by $\theta = kx + ly - \omega t$. At any given instant, the phase (the surface height) is constant on the lines of constant $kx + ly$. The wave properties are constant along the lines of constant phase for the plane wave. Substituting the expression for the wave (equation 4.156) into equation (4.27), and simplifying, leads to the relation

$$\omega\eta_0[f^2 - \omega^2 + C_0^2K^2] = 0 \tag{4.157}$$

© 2015 by Taylor & Francis Group, LLC

Here K represents the wave number and $C_0 = (gH_0)^{1/2}$ is the phase speed in shallow water theory.

For unsteady motions, it is obvious that $\omega \neq 0$, and if $\eta_0 \neq 0$, then the plane wave solution corresponding to the dispersion relation (equation 4.157) will be determined by the zeros of the relation for different values of K. For this simple case, the dispersion relation may be solved explicitly and is given by

$$\omega = \omega(K) = \pm[f^2 + C_0^2 K^2]^{1/2} \tag{4.158}$$

The solution for ω implies that it is a function of the absolute value of $\mathbf{K}$ and not on its orientation. For a given $\mathbf{K}$, two free oscillations (solutions) are present with the phase speed being given by

$$C = \pm \left[C_0^2 + \frac{f^2}{K^2}\right]^{1/2} \tag{4.159}$$

For the case when rotation is absent ($f = 0$), the phase speed of these waves for all wavelengths moves with $(gH_0)^{1/2}$. The presence of rotation increases the wave speed. An interesting conclusion from equation (4.158) is that all free waves have frequencies which exceed f, i.e., have periods less than half a rotational period.

Equation (4.156) can be written as

$$\eta = |\eta_0| \cos(kx + ly - \omega t + \phi) \tag{4.160}$$

In the above expression, if we assume that $\eta_0 = |\eta_0| e^{i\phi}$, then we can write the expression for velocity components u and v. In fact, if we denote them as $u_\parallel$ and $u_\perp$, respectively, the expression for them will be as follows:

$$u_\parallel = \frac{|\eta_0|}{H_0} C \cos(kx + ly - \omega t + \phi) \tag{4.161}$$

$$u_\perp = \frac{|\eta_0|}{H_0} \frac{f}{\omega} C \sin(kx + ly - \omega t + \phi) \tag{4.162}$$

where

$$u_\parallel = \left[\mathbf{u}_H \cdot \frac{\mathbf{K}}{K}\right] \tag{4.163}$$

$$u_\perp = [\mathbf{u}_H - u_\parallel \mathbf{K}] \tag{4.164}$$

Here $u_\parallel$ is the particle velocity parallel to $\mathbf{K}$ or perpendicular to the wave crests, whereas $u_\perp$ is the particle speed along the crests. The relation between the two expressions for the velocity components may be simplified to yield

$$u_\parallel^2 + u_\perp^2 \frac{\omega^2}{f^2} = \frac{C^2 \eta_0^2}{H_0^2} \tag{4.165}$$

© 2015 by Taylor & Francis Group, LLC

4.12 Poincaré and Kelvin Waves

In this section, we study the free linear modes of oscillation of a shallow, rotating fluid. The aim is to study the modes which appear in a partially bounded region, i.e., a channel of width L, oriented parallel to the x axis, the horizontal axis. Since the region is bounded in the vertical direction, y direction, the condition on the behavior of the flow at infinity need not be satisfied. However, if the channel is rigid, the velocity in the vertical direction must vanish (no slip condition), which means that

$$\frac{\partial^2 \eta}{\partial y \partial t} - f\frac{\partial \eta}{\partial x} = 0 \qquad y = 0, L \tag{4.166}$$

The governing wave equation for η, with the assumption that H_0 is a constant, is given by

$$\frac{\partial}{\partial t}\left[\left(\frac{\partial^2}{\partial t^2} + f^2\right)\eta - C_0^2 \boldsymbol{\nabla}^2 \eta\right] = 0 \tag{4.167}$$

We seek solutions which are periodic in x and t of the form

$$\eta = \Re\left[\hat{\eta}(y) e^{i(kx - \omega t)}\right] \tag{4.168}$$

where $\hat{\eta}$ is the (complex) amplitude of the wave and varies with respect to the coordinate y.

Substituting (4.168) into equation (4.166) leads to the eigenvalue problem for the amplitude, for which the differential equation is:

$$\frac{d^2 \hat{\eta}}{dy^2} + \left[\frac{\omega^2 - f^2}{C_0^2} - k^2\right]\hat{\eta} = 0 \tag{4.169}$$

with the boundary condition

$$\frac{d\hat{\eta}}{dy} + f\frac{k}{\omega}\hat{\eta} = 0 \qquad y = 0, L \tag{4.170}$$

The general solution of the eigenvalue problem may be written as

$$\hat{\eta} = A \sin \alpha y + B \cos \alpha y \tag{4.171}$$

where

$$\alpha^2 = \frac{\omega^2 - f^2}{C_0^2} - k^2 \tag{4.172}$$

It is important to determine the coefficients A and B. For this, we apply the

© 2015 by Taylor & Francis Group, LLC

boundary conditions, which will yield two homogeneous equations for A and B given by

$$\alpha A + \frac{fk}{\omega} B = 0 \quad (4.173)$$

$$A[\alpha \cos \alpha L + f\frac{k}{\omega} \sin \alpha L] + B[\frac{fk}{\omega} \cos \alpha L - \alpha \sin \alpha L] = 0 \quad (4.174)$$

In order to obtain nontrivial solutions for A and B, we equate the determinant of the coefficient matrix to be zero so that a relation between ω, k, and f results. This will be the dispersion relation as discussed in earlier sections. Simple algebra yields

$$(\omega^2 - f^2)(\omega^2 - C_0^2 k^2) \sin \alpha L = 0 \quad (4.175)$$

The above equation is the product of three terms. Thus, $\sin \alpha L = 0$ or $\omega^2 = f^2$ and finally $\omega^2 = C_0^2 k^2$. Let us consider all the three possibilities separately.

Case (i): The equation

$$\sin \alpha L = 0 \quad (4.176)$$

can be satisfied if α satisfies

$$\alpha = \frac{\pi n}{L}, \qquad n = 1, 2, 3, \cdots \quad (4.177)$$

The above relation implies an infinite number of solutions. It is important to note that $\alpha = n = 0$, which we did not specify above, is not a solution. Such a solution would be a plane wave with crests oriented parallel to the vertical y axis. Thus, there will be no variation in the y direction. However, for a nonrotating fluid, such a solution is possible. With the constraint that $\partial \eta / \partial y = 0$, a possible solution is given by

$$v = \frac{gf}{f^2 - \omega^2} \frac{\partial \eta}{\partial x}(x, t) \quad (4.178)$$

The above solution is completely different from zero and is independent of y, which means that it cannot satisfy the boundary condition of vanishing v at the boundaries $y = 0, L$. From the definition of α, we see that the condition for case (i) may be written as

$$\alpha^2 = \frac{\omega^2 - f^2}{C_0^2} - k^2 = 0 \quad (4.179)$$

In other words:

$$\omega = \omega_\pi = \pm \left[f^2 + C_0^2 \left(\frac{k^2 + n^2 \pi^2}{L^2} \right) \right]^{1/2}, \qquad n = 1, 2, 3, \cdots \quad (4.180)$$

The above relation describes the Poincaré waves which are similar dynamically

© 2015 by Taylor & Francis Group, LLC

to the plane wave oscillations described in the earlier section. An interesting observation is that the solution for the Poincaré waves has two modes, of the same magnitude but of opposite sign. This implies that Poincaré waves propagate with their phase velocity in both the positive and the negative horizontal x direction. Also their frequencies are always greater than f. Mathematically,

$$\omega \geq \left[f^2 + \frac{C_0^2 \pi^2}{L^2} \right]^{1/2} \tag{4.181}$$

The dynamical variables can be calculated from the relation between equations (4.171) and (4.178). For example, with simple algebra, we can determine η as

$$\eta = \eta_0 \left[\cos \frac{n\pi y}{L} - \frac{L}{n\pi} \frac{f}{C_x} \sin \frac{n\pi y}{L} \right] \cos(kx - \omega t + \phi) \tag{4.182}$$

where $C_x = \omega/k$ is the phase speed in the horizontal x direction, η_0 is an arbitrary amplitude, and ϕ is an arbitrary phase. In a similar way, the velocity components u and v may be calculated.

Case (ii): The second solution of the dispersion relation can be written as

$$\omega = \pm C_0 k \tag{4.183}$$

This solution is one of the most simple solutions of the dispersion relation. This is similar to the dispersion relation for a plane wave whose crests are parallel to the vertical y direction in a nonrotating fluid. This can be viewed as a supplement solution of the Poincaré waves with the fact that it plays a role of $n = 0$ mode. We made a statement that this is not valid for rotating fluid. However, the solution mentioned above is a solution of the dispersion relation and is referred to as the Kelvin mode. This is one of the most fundamental modes for rotating fluids. Let us look at the dynamical structure of this mode.

Consider the behavior of this mode in the propagation characteristics in the positive x direction, i.e., the solution $\omega = C_0 k$. Then,

$$\alpha^2 = -\frac{f^2}{C_0^2} \tag{4.184}$$

This makes α purely imaginary, i.e., $\alpha = \pm if/C_0$. If we assume without loss of generality that $\alpha = if/C_0$, then we can write the dynamical variables as

$$\eta = \eta_0 e^{-fy/C_0} \cos(k[x - C_0 t] + \phi) \tag{4.185}$$

$$u = \frac{\eta_0}{H_0} C_0 e^{-fy/C_0} \cos(k[x - C_0 t] + \phi) = -\frac{g}{f} \frac{\partial \eta}{\partial y} \tag{4.186}$$

$$v = 0 \tag{4.187}$$

Equations (4.185 and 4.186) can be simplified to yield the classical wave equation as

$$\frac{\partial^2 \eta}{\partial t^2} = C_0^2 \frac{\partial^2 \eta}{\partial x^2} \tag{4.188}$$

© 2015 by Taylor & Francis Group, LLC

The e-folding scale for the channel amplitude may be defined as

$$R = \frac{C_0}{f}$$

The physical interpretation of the above expression is that to render the free surface flat the distance over which the gravitational tendency (C_0) is balanced by the Coriolis acceleration (f) to deform the surface. In the limit $R \to \infty$ as $f \to 0$, the Kelvin wave becomes in the limit the $n = 0$ mode. In this way, one can say that the Kelvin mode is a limiting case of the Poincaré mode. For $\omega = -C_0 k$, i.e., for a Kelvin wave propagating in the negative direction has similar behavior.

The amplitude of the Kelvin wave diminishes exponentially with the maximum at the wall, $y = L$, so that the wave height is maximum on the right side of the observer looking in the direction of the wave propagation. For the Kelvin wave to exist, it is necessary to have at least one internal boundary. This is because for a completely infinite region, the exponential increase of the wave is not allowed. Also for a low frequency Kelvin wave, for which $\omega \ll f$, it is important that the condition $kR \ll 1$ also holds. This means that the wavelength in the horizontal direction is much larger compared to the cross stream scale R.

These waves are essentially ones whose frequency ω is same as the Coriolis frequency f. The frequency of these waves is called the inertial frequency, i.e., an oscillation for which the frequency has two solutions related to the Coriolis frequency. The starting point for the discussion is the dispersion relation mentioned in equation (4.175). We had discussed in detail two solutions of the possible three roots of the dispersion relation. Let us return to the third root and discuss its characteristics. The third root of the dispersion relation (4.175) is

$$\omega = \pm f \tag{4.189}$$

For case (iii) when $\omega = f$, the operator $\partial^2/\partial t^2 + f^2$ must have a unique inverse so that the velocity component v can be determined in terms of η. Care must be taken so that this is not a spurious root. If we assume the Fourier description in equations (4.18−4.20), then with the assumption that $\omega = f$, these equations reduce to

$$f\hat{u} - if\hat{v} = gk\hat{\eta} \tag{4.190}$$

$$f\hat{u} - if\hat{v} = -g\frac{d\hat{\eta}}{dy} \tag{4.191}$$

where $\hat{u}$ and $\hat{v}$ are the complex amplitudes of u and v:

$$u = \Re\left[\hat{u}e^{i(kx-ft)}\right] \tag{4.192}$$

$$v = \Re\left[\hat{v}e^{i(kx-ft)}\right] \tag{4.193}$$

© 2015 by Taylor & Francis Group, LLC

The additional constraint to solve the equation is given by

$$\frac{d\hat{\eta}}{dy} + k\hat{\eta} = 0 \tag{4.194}$$

so that

$$\hat{\eta} = \eta_0 e^{-ky} \tag{4.195}$$

The equation for $\hat{v}$ can be derived to yield

$$\frac{d\hat{v}}{dy} - k\hat{v} = \frac{if\hat{\eta}}{H_0}\left(1 - \frac{k^2 C_0^2}{f^2}\right) \tag{4.196}$$

The general solution of the above equation can be written as

$$\hat{v} = V_0 e^{ky} - \frac{if}{2k}\frac{\eta_0}{H_0}(1 - k^2R^2)e^{-ky} \tag{4.197}$$

where V_0 is an arbitrary constant. However, one has to impose the condition that $\hat{v}(0)$ vanishes. This imposes a restriction on V_0. Applying the boundary condition and simplifying lead to the solution for $\hat{v}$ as

$$\hat{v} = \frac{if\eta_0}{kH_0}(1 - k^2R^2)\sinh ky \tag{4.198}$$

One can easily check that the condition that $\hat{v}$ vanishes at $y = L$ is possible only for the special case $k = R^{-1}$, in which v vanishes identically. The oscillation $\omega = f$ is a spurious root of the eigenvalue problem. Thus, the complete spectrum consists of the Kelvin mode, the Poincaré mode, and the $\omega = 0$ mode.

4.13 Lamb and Rayleigh Waves

We shall discuss first Lamb waves. They are very important and play a crucial role in solid mechanics. For the sake of completeness, we shall briefly mention their importance in solid mechanics, after a discussion of Lamb waves in hydrodynamics.

The linearized equations of motion with the assumption that $\delta u = \hat{u}(z) \cdot exp[i(kx + \omega t)]$, $\delta v = \hat{v} \cdot exp[i(kx + \omega)]$ etc., are

$$i\omega\hat{u} - f\hat{v} + ik\frac{\hat{p}}{\rho_0} = 0 \tag{4.199}$$

$$i\omega\hat{v} + f\hat{u} = 0 \tag{4.200}$$

$$n_4 i\omega\hat{w} + \frac{d}{dz}[\frac{\hat{p}}{\rho_0} - n_3 B\frac{\hat{p}}{\rho_0} - g\hat{\Theta} = 0 \tag{4.201}$$

$$n_2 i\omega\frac{\hat{\rho}}{\rho_0} + ik\hat{u} + \frac{d}{dz}\hat{w} - \frac{n_1}{H_0}\hat{w} = 0 \tag{4.202}$$

$$i\omega\hat{\Theta} + B\hat{w} = 0 \tag{4.203}$$

© 2015 by Taylor & Francis Group, LLC

Here H_0, Θ, B, and n are scale height, potential temperature, static stability parameter, and amplitude parameters, respectively. The above equations are independent of x and t. The operators $\partial\partial x$ and $\partial/\partial t$ have been replaced by ik and $i\omega$, respectively,

Equations (4.199) and (4.200) can be simplified to yield

$$\hat{u} = -\frac{\omega k}{\omega^2 - f^2}\frac{\hat{p}}{\rho_0} \tag{4.204}$$

$$\hat{v} = -\frac{ifk}{\omega^2 - f^2}\frac{\hat{p}}{\rho_0} \tag{4.205}$$

Replacing $\hat{u}$ from equation (4.204) into (4.202) and using (4.203), we obtain

$$\hat{\Theta} = \frac{1}{\gamma}\frac{\hat{p}}{p_0} - \frac{\hat{\rho}}{\rho_0} = \frac{1}{C^2}\frac{\hat{p}}{\rho_0} - \frac{\hat{\rho}}{\rho_0} \tag{4.206}$$

where $C = \sqrt{\gamma R T_0}$ is the speed of sound. Again equations (4.202) and (4.203) reduce to

$$\frac{d}{dz}\hat{w} + \left[Bn_2 - \frac{n_1}{H_0}\right]\hat{w} + i\omega\left[\frac{n^2}{C^2} - \frac{k^2}{\omega^2 - f^2}\right]\frac{\hat{p}}{\rho_0} = 0 \tag{4.207}$$

$$i\omega\frac{d}{dz}\frac{\hat{p}}{\rho_0} - i\omega B n_3\frac{\hat{p}}{\rho_0} + (gB - n_4\omega^2)\hat{w} = 0 \tag{4.208}$$

The final second-order differential equation for $\hat{w}$, by simplification, yields

$$\omega\frac{d^2}{dz^2}\hat{w}(z) + \left[B(n_2 - n_3) - \frac{n_1}{H_0}\right]\frac{d}{dz}\hat{w}(z) + (gB - n_4\omega^2)\left[\frac{k^2}{\omega^2 - f^2} - \frac{n_2}{C^2}\right] - Bn_3\left[Bn_2 - \frac{n_2}{H_0}\right]\hat{w}(z) = 0 \tag{4.209}$$

Let us discuss the solution of the above equation. One of the simplest solutions is $\omega = 0$. This obviously does not exhibit a wave. The other solution (trivial), $\hat{w} = 0$ for all values of z, is defined as the Lamb wave. The case $\omega \neq 0$ and $\hat{w} \neq 0$ for all z needs a solution of the differential equation mentioned above.

Introducing the transformation

$$\hat{w}(z) = \bar{w}(z)\exp\left[-\frac{1}{2}\left\{B(n_2 - n_3) - \frac{n_1}{H_0}\right\}z\right]$$

leads to a simpler equation where the first-order derivatives are not present as

$$\frac{d^2\bar{w}}{dz^2}\left[(gB - n_4\omega^2)\left[\frac{k^2}{\omega^2 - f^2} - \frac{n_2}{C^2}\right]\right] - Bn_3\left[Bn_2 - \frac{n_1}{H_0}\right] - \frac{1}{4}\left[B(n_2 - n_3) - \frac{n_1}{h_0}\right]^2\bar{w} = 0 \tag{4.210}$$

© 2015 by Taylor & Francis Group, LLC

$$m^2 = \left[(gB - n_4\omega^2)\left[\frac{k^2}{\omega^2 - f^2} - \frac{n_2}{C^2}\right]\right] - Bn_3\left[Bn_2 - \frac{n_1}{H_0}\right] - \frac{1}{4}\left[B(n_2 - n_3) - \frac{n_1}{h_0}\right]^2 \quad (4.211)$$

The above relation is of fourth order in the frequency ω. The solution for $\bar{w}$ which is an ordinary differential equation with constant coefficients can be written as $\bar{w} \propto (imz)$.

Let us briefly discuss Rayleigh and Lamb waves and their importance in solid mechanics. It is important to note that the wave equations for Rayleigh waves are the same as those for bulk waves. The one difference is that the boundary condition for Rayleigh waves is slightly different from that of bulk waves. These waves whose properties are determined by the presence of boundaries are often referred to as guided waves. The Rayleigh solution for a medium with a single boundary is the simplest of such waves. For more complicated boundaries, solutions are tried using numerical methods. However, the simplest geometry is that of a plate which we will discuss now.

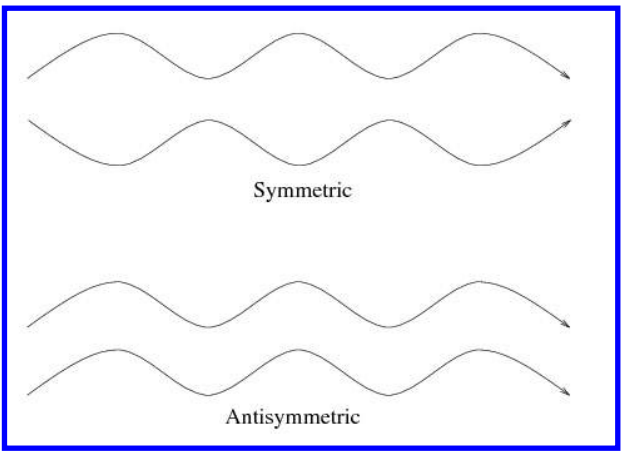


FIGURE 4.4: An example of the symmetric and anti-symmetric Lamb modes.

The initial analysis for thin plates was initiated by Lamb and in recognition of this, these solutions are termed the Lamb waves. Consider a geometry in which a solid medium bounded by two parallel plates which is at a distance of $2d$ apart. The horizontal direction is the x direction, which we assume as the direction of propagation. There are two types of solutions. The first solution is an even function, i.e., $u(z) = u(-z)$. Such solutions are the symmetric solutions. The other solution, termed anti-symmetric, is such that u is an odd function, i.e., $u(-z) = -u(z)$. Examples of Lamb waves are given in Figure 4.4.

The characteristic equations for the symmetric and antisymmetric modes

© 2015 by Taylor & Francis Group, LLC

are given by

$$\frac{\tan(qd)}{\tan(pd)} = -\frac{4k^2pq}{(q^2-k^2)^2} \tag{4.212}$$

$$\frac{\tan(pd)}{\tan(qd)} = -\frac{4k^2pq}{(q^2-k^2)^2} \tag{4.213}$$

p and q are, respectively, given by $p = \frac{\omega}{\sqrt{C_L^2 - C^2}}$, and $q = \frac{\omega}{\sqrt{C_T^2 - C^2}}$.

4.14 Inertial Waves

Inertial waves have to do with waves in a rotating fluid. We assume that the fluid is incompressible and rotating uniformly. Waves are generated due to Coriolis forces acting on the fluid due to rotation. Consider the fluid in rotating coordinate geometry. The mechanical equations will include contributions from centrifugal and Coriolis terms. The forces (per unit mass of fluid) should be added to the right-hand side of Euler's equations of motion. The centrifugal force is written as $\boldsymbol{\nabla}(1/2)(\boldsymbol{\Omega}\times\mathbf{r}^2)$, where $\boldsymbol{\Omega}$ is the angular velocity of the fluid. The effective pressure term can be written as

$$P = p - \frac{1}{2}\rho(\boldsymbol{\Omega}\times\mathbf{r})^{\mathbf{2}} \tag{4.214}$$

The Coriolis force acts only when the fluid has a motion relative to the rotating coordinates and is given by $2\mathbf{v}\times\boldsymbol{\Omega}$, $\mathbf{v}$ being the velocity in those coordinates. Euler's equations of motions are

$$\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\boldsymbol{\nabla})\mathbf{v} + \mathbf{2\Omega}\times\mathbf{v} = -\frac{\mathbf{1}}{\rho}\boldsymbol{\nabla}\mathbf{P} \tag{4.215}$$

For the incompressible fluid, the equation of continuity simplifies to $\boldsymbol{\nabla}\cdot\mathbf{v} = 0$. We shall assume that the wave amplitude is small, so that the quadratic terms can be neglected. The above equation reduces to

$$\frac{\partial\mathbf{v}}{\partial t} + 2\boldsymbol{\Omega}\times\mathbf{v} = -\left(\frac{1}{\rho}\right)\boldsymbol{\nabla}p' \tag{4.216}$$

where p' denotes the variation in the pressure, while ρ is a constant. Taking curl on both sides of the above equation, the pressure term can be eliminated to yield

$$\boldsymbol{\nabla}\times(\boldsymbol{\Omega}\times\mathbf{r}) = \boldsymbol{\Omega}\boldsymbol{\nabla}\cdot\mathbf{v} - (\boldsymbol{\Omega}\cdot\boldsymbol{\nabla})\mathbf{v} \tag{4.217}$$

$$= -(\boldsymbol{\Omega}\cdot\boldsymbol{\nabla})\mathbf{v} \tag{4.218}$$

© 2015 by Taylor & Francis Group, LLC

Taking the direction of $\mathbf{\Omega}$ as the direction of the z axis, we can simplify the above equation as

$$\frac{\partial}{\partial t}\boldsymbol{\nabla}\times\mathbf{v} = 2\mathbf{\Omega}\frac{\partial \mathbf{v}}{\partial z} \tag{4.219}$$

Plane wave solutions of the form

$$\mathbf{v} = \mathbf{A}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega\mathbf{t})} \tag{4.220}$$

satisfy the relation

$$\mathbf{k}\cdot\mathbf{A} = 0 \tag{4.221}$$

Substituting equation (4.220) into equation (4.219) yields

$$\omega\mathbf{k}\times\mathbf{v} = 2i\Omega k_z\mathbf{v} \tag{4.222}$$

Vector multiplication by $\mathbf{k}$ and simplification yields the dispersion relation as

$$-\omega k^2\mathbf{v} = 2i\Omega_z\mathbf{k}\times\mathbf{v} \tag{4.223}$$

The relation between ω and $\mathbf{k}$ is given by

$$\omega = 2\Omega k_z/k = 2\Omega\cos\theta \tag{4.224}$$

where θ is the angle between $\mathbf{k}$ and $\mathbf{\Omega}$. If we assume that the amplitude of the wave is complex and takes the form $\mathbf{A} = \mathbf{a} + \mathbf{ib}$, with real vectors $\mathbf{a}$ and $\mathbf{b}$, it follows that $\mathbf{n}\times\mathbf{b} = \mathbf{a}$. Separating the real and imaginary parts, we obtain

$$v_x = a\cos(\omega t - \mathbf{k}\cdot\mathbf{r}), \qquad v_y = -a\sin(\omega t - \mathbf{k}\cdot\mathbf{r}) \tag{4.225}$$

The above expression indicates that the wave has a circular polarization, i.e., at each point in the space, the vector $\mathbf{v}$ rotates in the course of time while the magnitude remains constant. The velocity of wave propagation is given by

$$\mathbf{U} = \frac{\partial\omega}{\partial\mathbf{k}} = \frac{2\Omega}{k}[\nu - \mathbf{n}(\mathbf{n}\cdot\nu)] \tag{4.226}$$

in which ν may be considered as a unit vector along the rotation. This is the case for internal gravity waves. Their magnitude and component along $\mathbf{\Omega}$ are

$$U = \frac{2\Omega}{k}\sin\theta, \qquad \mathbf{U}\cdot\nu = \frac{\mathbf{2\Omega}}{\mathbf{k}}\sin^{\mathbf{2}}\theta = \mathbf{U}\sin\theta \tag{4.227}$$

The above modes of the dispersion relation are called inertial waves. The energy of the waves are entirely the kinetic energy, as it is well known that the Coriolis forces do no work on the moving fluid. A specific case of the axially symmetric inertial waves propagates along the axis of rotation of the fluid. We shall discuss this example briefly.

© 2015 by Taylor & Francis Group, LLC

4.14.1 Axisymmetric Waves

We shall discuss the case of an axially symmetric wave propagating along the axis of an incompressible, rotating (uniform) fluid as a whole. We shall take the cylindrical polar coordinates r, ϕ, z with the z axis parallel to $\mathbf{\Omega}$. Taking the components of the linearized equations of motion we have

$$-i\omega v_r - 2\Omega v_\phi = -\frac{1}{\rho}\frac{\partial p'}{\partial r} \tag{4.228}$$

$$-i\omega v_\phi + 2\Omega v_r = 0, \qquad -i\omega v_z = -\left(\frac{ik}{\rho}\right) p' \tag{4.229}$$

The above equations have to be combined with the mass conservation equation. Define the transformation

$$v_r = G(r)e^{i(\omega t - kx)}$$

which yields a single equation for G as

$$\frac{d^2G}{dr^2} + \frac{1}{r}\frac{dF}{dr} + \left\{\frac{4\Omega^2k^2}{\omega^2} - k^2 - \frac{1}{r^2}\right\} G = 0 \tag{4.230}$$

The solution for the above equation can be written in terms of Bessel functions J_1 of order unity. An interesting observation is that ω is independent of k. The possible values of the frequency depend on the condition $\omega < 2\Omega$. The above equation does not have a solution if this condition is not satisfied. For a rotating fluid, bounded by a cylindrical wall of radius R, the condition $v_r = 0$ at the wall yields the relation

$$ka\sqrt{[(4\Omega^2/\omega^2) - 1]} = x_n$$

between ω, the frequency, and the given n.

4.15 Rossby Waves

In this section, we shall consider free oscillations in a channel of width L with the assumption that the height H_0 varies slightly in the y direction as follows:

$$h_0 = D_0\left(1 - \frac{sy}{L}\right) \tag{4.231}$$

where the slope $s \ll 1$.

An important aspect of these waves is that if H_0 are constant along the horizontal axis, this lead to pure geostrophic motions only if v is identically

© 2015 by Taylor & Francis Group, LLC

zero. Equation (4.27) will be the starting point of further analysis. Assume wave solutions of the form

$$\eta = \Re\left\{\hat{\eta}(y)e^{i(kx-\omega t)}\right\} \tag{4.232}$$

yields an equation for $\hat{\eta}$ as follows:

$$\left\{1 - s\frac{y}{L}\right\}\frac{d^2\hat{\eta}}{dy^2} - \frac{s}{L}\frac{d\hat{\eta}}{dy} + \hat{\eta}\left[\frac{\omega^2 - f^2}{gD_0} - k^2\left(1 - s\frac{y}{L}\right) - \frac{fs}{L\omega}k\right] = 0 \tag{4.233}$$

with the constraint that

$$\frac{d\hat{\eta}}{dy} + \frac{fk}{\omega}\hat{\eta} = 0 \tag{4.234}$$

Assuming that y/L is very small compared to unity, for small s, an approximate equation can be written as

$$\frac{d^2\hat{\eta}}{dy^2} - \frac{s}{L}\frac{d\hat{\eta}}{dy} + \hat{\eta}\left[\frac{\omega^2 - f^2}{C_0^2} - k^2 - \frac{f}{L}\frac{s}{\omega}k\right] = 0 \tag{4.235}$$

The solution of the above equation can be written as

$$\hat{\eta} = e^{ty/2L}[A\sin\alpha y + B\cos\alpha y] \tag{4.236}$$

where the expression for α is as follows:

$$\alpha^2 = \frac{\omega^2 - f^2}{C_0^2} - \left\{k^2 + \frac{s^2}{4L^2}\right\} - \frac{fks}{\omega L} \tag{4.237}$$

The eigenvalue problem can be written as

$$(\omega^2 - f^2)(\omega^2 - k^2C_0^2)\sin\alpha L = 0 \tag{4.238}$$

The solution corresponding to the zeros of $\sin\alpha L$ yields

$$\omega^2 - \frac{fksC_0^2}{L\omega} - C_0^2\left(k^2 + \frac{n^2\pi^2}{L^2} + \frac{f^2}{C_0}\right) = 0 \tag{4.239}$$

The term $O(s^2/L^2)$ has been ignored with respect to $n^2\pi^2/L^2$ since we have assumed $s \ll 1$. Also n must be greater than zero. The first class of solutions has frequencies which exceed f. For this class, the term in s may be neglected and to the $O(s)$, one can recover the Poincaré modes, which are

$$\omega^2 = f^2 + C_0^2\left(k^2 + \frac{n^2\pi^2}{L^2}\right) + O(s), \qquad n = 1, 2, 3, \cdots \tag{4.240}$$

The high-frequency Poincaré modes are unaffected by the presence of a minor change in the slope of the bottom. The new solution, which is of the order $O(1)$, leads to the dispersion relation of the Rossby waves, namely,

$$\omega = -s\left(\frac{f}{L}\right)\frac{k}{k^2 + n^2\pi^2/L^2 + f^2/C_0^2}, \qquad n = 1, 2, 3, \cdots \tag{4.241}$$

© 2015 by Taylor & Francis Group, LLC

The maximum value of the Rossby wave frequency is attained when

$$k = k_n = \left\{ \frac{n^2\pi^2}{L^2} + \frac{f^2}{C_0^2} \right\}^{1/2} \tag{4.242}$$

for which

$$\omega = \omega_{max} = -\frac{s}{2} \frac{f}{(n^2\pi^2 + f^2L^2/C_0^2)^{1/2}} \tag{4.243}$$

Thus, for small s, the Rossby wave frequency is always less than f. The Rossby wave which depends on both f and s should be nonzero. Its rotation period is greater than the rotation period (as the frequency is less than f). An important observation about the phase velocity of the Rossby wave is that it is always negative in the x direction as is clear from the relation below:

$$C_s = \frac{\omega}{k} = -\left\{ \frac{sf}{L} \right\} / \left\{ k^2 + \frac{n^2\pi^2}{L^2} + \frac{f^2}{C_0^2} \right\} \tag{4.244}$$

The dynamical fields for the Rossby waves are given by

$$\eta = \eta_0 \sin \frac{\pi n y}{L} \cos(kx - \omega t + \phi) + O(s) \tag{4.245}$$

$$u = -\frac{g}{f} \frac{n\pi}{L} \eta_0 \cos \frac{n\pi y}{L} \cos(kx - \omega t + \phi) + O(s) \tag{4.246}$$

$$u = -\frac{g}{f} k \eta_0 \sin \frac{n\pi y}{L} \sin(kx - \omega t + \phi) + O(s) \tag{4.247}$$

4.16 Forced Stationary Waves in the Atmosphere

In this section, we discuss waves with an application to the Earth's atmosphere. Consider a zonal flow, which is independent of the longitude coordinate as follows:

$$u = u_0(y, z) \tag{4.248}$$

with the corresponding stream function

$$\psi = \psi_0(y, z) = -\int_0^y u_0(y', z) dy' \tag{4.249}$$

and a heating term

$$\mathcal{H} = \mathcal{H}(x, y). \tag{4.250}$$

Define a new stream function as follows:

$$\psi = \psi_0(y, z) + \phi(x, y, z) \tag{4.251}$$

© 2015 by Taylor & Francis Group, LLC

where $\phi(x, y, z)$ is a disturbance (assumed to be small) produced by the heating and topography.

Substituting the above expression into the vorticity equation and retaining linear terms, the following equation results:

$$u_0 \frac{\partial}{\partial x}\left\{\frac{1}{\rho_s}\frac{\partial}{\partial z}\frac{\rho_s}{S}\frac{\partial \phi}{\partial z} + \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2}\right\} + \left\{\beta - \frac{\partial^2 u_0}{\partial y^2} - \frac{1}{\rho_s}\frac{\partial}{\partial z}\frac{\rho_s}{S}\frac{\partial u_0}{\partial z}\right\}\frac{\partial \phi}{\partial x} = \frac{1}{\rho_s}\frac{\partial}{\partial z}\left\{\frac{\rho_s \mathcal{H}}{S}\right\} \quad (4.252)$$

The solution in terms of

$$\phi = \Re\left\{\Phi(z) e^{ikx} \sin \pi y\right\}$$

is sought subject to Φ satisfying the following conditions:

$$\frac{1}{\rho_s}\frac{d}{dz}\frac{\rho_s}{S}\frac{d\Phi}{dz} + \left\{\frac{\beta}{u_0} - K^2\right\}\Phi = \frac{1}{\rho_s}\frac{\partial}{\partial z}\frac{\rho_0 \mathcal{H} e^{-\alpha x} e^{i\theta_H}}{iku_0} \quad (4.253)$$

and

$$S^{-1}\frac{d\Phi}{dz} = -\eta_0 + S^{-1}\frac{\mathcal{H} e^{i\theta_H}}{iku_0} \quad (4.254)$$

where $K^2 = k^2 + \pi^2$. For large values of z, it is important that $\rho_s|\Phi|^2$ remains finite as $z \to \infty$. Let us consider the simpler case of S being constant with the nondimensional scale height H defined as

$$H = -\left[\frac{1}{\rho_s}\frac{\partial \rho_s}{\partial z}\right] \quad (4.255)$$

which is also a constant. Equation (4.253) simplifies to

$$\frac{d^2\Phi}{dz^2} - \frac{1}{H}\frac{d\Phi}{dz} + \left\{\frac{\beta}{u_0} - K^2\right\} S\Phi = -\frac{(1 + (\alpha H)^{-1})\alpha \mathcal{H} e^{-\alpha x + i\theta_h}}{iku_0} \quad (4.256)$$

The homogeneous solutions of the above equation are given by

$$\Phi_H(Z) = e^{z/2H}[Ae^{imz} + Be^{-imz}] \quad (4.257)$$

where

$$m = \left\{\left(\frac{\beta}{u_0} - K^2\right) S - \frac{1}{4H^2}\right\}^{1/2}$$

The nature of the solutions depends on whether

$$\frac{\beta}{u_0} > K^2 + (4H^2 S)^{-1}$$

is positive or negative. For $m^2 > 0$, the solution will be oscillatory in z, while for $m^2 < 0$, the solution is either growing or decaying exponentially, without the oscillation.

© 2015 by Taylor & Francis Group, LLC

4.17 Solitary Waves − KdV Equation

In this section, we derive the nonlinear evolution equation which is the famous Kortweg−deVries equation in a two-layer system. The basic equations of motion in a rotating frame are given by

$$\frac{D\mathbf{u}}{Dt} + 2\boldsymbol{\Omega}\times\mathbf{u} = -\frac{1}{\rho}\boldsymbol{\nabla}p + B\hat{z} \tag{4.258}$$

$$\boldsymbol{\nabla}\cdot\mathbf{u} = 0 \tag{4.259}$$

$$\frac{DB}{Dt} + N^2 w = 0 \tag{4.260}$$

where B is a static stability parameter.

We have two assumptions, namely, the effect of Coriolis force is neglected and that $\partial/\partial y = 0$. In the two layer system, both the layers are neutrally stratified, i.e. $N = 0$, where N is the Brunt–Vaisala frequency. With $N = 0$, $B = 0$ is automatically satisfied. The basic equations in component form are written as

$$u_{i,t} + u_i u_{i,x} + w_i u_{i,z} = -\frac{1}{\rho}p_{i,x} \tag{4.261}$$

$$w_{i,t} + u_i w_{i,x} + w_i w_{i,z} = -\frac{1}{\rho}p_{i,z} \tag{4.262}$$

$$u_{i,x} + w_{i,z} = 0 \tag{4.263}$$

Here, the index takes the value 1 or 2, depending on the upper or lower layer, respectively. The subscript, x, z, or t, denotes derivative with respective to the independent variables. A simple model for the solitary wave is given in Figure 4.5.

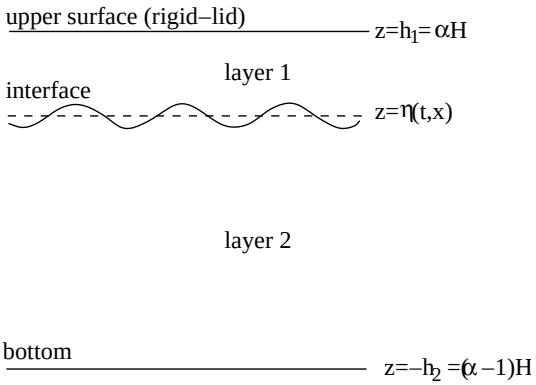


FIGURE 4.5: Two-layer model for a solitary wave.

© 2015 by Taylor & Francis Group, LLC

We choose the z axis in such a way that the interface is at $z = 0$. The upper surface is denoted by $z = h_1$, while the lower surface is denoted by $z = -h_2$. We introduce a new scale H, such that $h_1 = \alpha H$ and $h_2 = (1 - \alpha)H$, which implies that $H = h_1 + h_2$ is the total depth of the water. The boundary conditions are defined as

$$w_1|_{z=\alpha H} = 0$$

$$w_2|_{z=(\alpha-1)H} = 0$$

The condition for the pressure is

$$p = p_0(z) + p'(t, \bar{x})$$

where p_0 is the hydrostatic pressure and the Brunt–Vaisala frequency is defined by

$$N^2 = -\frac{g}{\rho}\left(\frac{d\rho_0}{dz} + \frac{\rho_0 g}{C_s^2}\right)$$

Integrating the pressure in both layers yields

$$p_{0,1} = g\int_z^{h1} d\zeta \rho_{0,1}(\zeta), \qquad p_{0,2} = g\int_0^{h1} d\zeta \rho_{0,1}(\zeta) + g\int_z^0 d\zeta \rho_{0,2}(\zeta)$$

The constraint that the total pressure be continuous at the interface results in

$$p_2'|_{z=\eta} - p_1'|_{z=\eta} = g\int_0^{\eta} d\zeta \rho_{0,2}(\zeta) - g\int_0^{\eta} d\zeta \rho_{0,1}(\zeta)$$

Introducing the following scales:

$$[u_i] = \epsilon C, \quad [t] = T, \quad [w_i] = \epsilon CH/L, \quad [p_i'] = \rho\epsilon C^2, \quad [\eta] = \epsilon H$$

the equations of motion can be written as

$$u_{i,t} + \epsilon u_i u_{i,x} + \epsilon w_i u_{i,z} = -p_{i,x}' \tag{4.264}$$

$$\delta(w_{i,t} + \epsilon u_i w_{i,x} + \epsilon w_i w_{i,z}) = -p_{i,z}' \tag{4.265}$$

$$u_{i,x} + w_{i,z} = 0 \tag{4.266}$$

The new parameter δ is defined as $\delta = (H/L)^2$, which is a measure of the dispersion from the hydrostatic approximation. In the limit $\delta \to 0$, we recover the long-wave limit for which the waves are nondispersive. The boundary condition in nondimensional form has the following form:

$$w_1|_{z=\alpha} = 0, \quad w_2|_{z=\alpha-1} = 0$$

$$w_i|_{z=\epsilon\eta} = \eta_t + \epsilon\eta_x u_i|_{z=\epsilon\eta}$$

$$p_2'|_{z=\epsilon\eta} - p_1'|_{z=\epsilon\eta} = \eta$$

The phase speed $C^2 = g'H$, where g' is the reduced gravity is given by $g' =$

© 2015 by Taylor & Francis Group, LLC

$g(\rho_{0,2} - \rho_{0,1})/\rho$. In order to derive the KdV equation, we make the following assumption:

$$\epsilon = \mathcal{O}(\delta) \ll 1$$

The basic equation with the relevant boundary conditions is used to derive one single equation for the interfacial displacement η. We look for an equation of the form

$$D_0(\eta) + \epsilon D_1(\eta) + \epsilon^2 D_2(\eta) + \cdots = 0 \tag{4.267}$$

where D_i are differential operators. We expand the variables in terms of a series in ϵ as

$$p_i' = p_0^{(0)} + \epsilon p_i^{(1)} + \epsilon^2 p_i^{(2)} + \cdots$$

with the constraint that $p_i^n = \mathcal{O}(1)$. At the lowest order ϵ^0, the equations become

$$u_{i,t}^{(0)} + p_{i,x}^{(0)} = 0 \tag{4.268}$$

$$p_{i,z}^{(0)} = 0 \tag{4.269}$$

$$u_{i,x}^{(0)} + w_{i,z}^{(0)} = 0 \tag{4.270}$$

It is interesting to note that $p_i^{(0)}$ and $u_i^{(0)}$ are independent of z. The boundary conditions reduce to

$$w_1^{(0)}|_{z=\alpha} = 0, \quad w_2^{(0)}|_{z=\alpha-1} = 0, \quad w_i^{(0)}|_{z=0} = \eta_t^{(0)}$$

$$p_2^{(0)} - p_1^{(0)} = \eta^{(0)}$$

The horizontal momentum equation and boundary condition for pressure may be combined to yield

$$\bar{u}_t^{(0)} + \eta_x^{(0)} = 0 \tag{4.271}$$

Integration of the continuity equation over the upper and lower layers, respectively, gives

$$\eta_t^{(0)} - \alpha u_{1,x}^{(0)} = 0$$

$$\eta_t^{(0)} + (1-\alpha) u_{2,x}^{(0)} = 0$$

Multiplying the first equation by $(1-\alpha)$ and the second by α and adding results in

$$\eta_t^{(0)} + \nu^2 \bar{u}_x^{(0)} = 0 \tag{4.272}$$

where $\nu^2 = \alpha(1-\alpha)$.

Combining equations (4.271 and 4.272), yields

$$\eta_{tt}^{(0)} - \nu^2 \eta_{xx}^{(0)} = 0$$

The above equation describes linear long waves, whose solution can be written as $F(x+\nu t) + G(x-\nu t)$. The lowest order operator D_0 is given by

$$D_0(\eta) = \eta_t + \nu \eta_x \tag{4.273}$$

© 2015 by Taylor & Francis Group, LLC

Let us now proceed to the next order analysis. For the order $\epsilon^{(1)}$, the equations and the boundary conditions reduce to

$$u_{i,t}^{(1)} + u_i^{(0)} u_{i,x}^{(0)} = -p_{i,x}^{(1)}$$

$$w_{i,t}^{(0)} = -p_{i,z}^{(1)}$$

$$u_{i,x}^{(1)} + w_{i,z}^{(1)} = 0$$

The boundary conditions are

$$w_1^{(1)}|_{z=\alpha} = 0, \quad w_2^{(1)}|_{z=\alpha-1} = 0, \quad w_i^{(1)}|_{z=0} = \eta_t^{(1)} + (u_i^{(0)}\eta^{(0)})_x$$

$$p_2^{(1)}|_{z=0} - p_1^{(1)}|_{z=0} = \eta^{(1)}$$

From the horizontal momentum equation, with the boundary conditions for the pressure, we arrive at

$$\bar{u}_t^{(1)} + (2\alpha - 1)\bar{u}^{(0)}\bar{u}_x^{(0)} + \eta_x^{(1)} = 0 \tag{4.274}$$

Integrating the continuity equation over the upper and lower layers as mentioned above, we obtain

$$\eta_t^{(1)} + (u_1^{(0)}\eta^{(0)})_x - \alpha u_{1,x}^{(1)}|_{z=0} - (1/3)\delta\alpha^2\eta_{xxt}^{(0)} = 0$$

$$\eta_t^{(1)} + (u_2^{(0)}\eta^{(0)})_x + (1-\alpha)u_{2,x}^{(1)}|_{z=0} - (1/3)\delta(1-\alpha)^2\eta_{xxt}^{(0)} = 0$$

Simple algebra yields

$$\eta_{tt}^{(1)} - \nu^2\eta_{xx}^{(1)} + 3(1-2\alpha)(\eta^{(0)}\eta_x^{(0)})_x - (1/3)\delta\nu^4\eta_{xxxx}^{(0)} = 0 \tag{4.275}$$

Extracting the operation $\partial/\partial t - \nu\partial/\partial x$ from the above equation results in

$$\eta_t^{(1)} + \nu\eta_x^{(1)} + (3/2\nu)(2\alpha - 1)\eta^{(0)}\eta_x^{(0)} + (1/6)\delta\nu^3\eta_{xxx}^{(0)} = 0 \tag{4.276}$$

Comparing the above equation with the operator D_1, we have

$$D_1(\eta) = (3/2\nu)(2\alpha - 1)\eta\eta_x + (1/6)\delta\nu^3\eta_{xxx} \tag{4.277}$$

Combining the operators D_0 and D_1, we have

$$\eta_t + \nu\eta_x + (3/2\nu)\epsilon(2\alpha - 1)\eta\eta_x + (1/6)\delta\nu^3\eta_{xxx} + \cdots = 0 \tag{4.278}$$

The first four terms (with the higher order terms $(\cdots)$ ignored) in the above equation is the famous KdV equation. We shall give an example of the solution of the KdV equation in the next few lines.

© 2015 by Taylor & Francis Group, LLC

4.17.1 *KdV* Equation

Let us consider a simpler form of the KdV equation given by

$$4u_t = 6uu_x + u_{xxx}$$

We seek a solution $u(x,t)$ which is an arbitrary function of $x - ct$. Consider the form of the solution as

$$u(x,t) = U(x - ct)$$

A simple algebra yields the form of the equation for U as

$$-4cU' = 6UU' + U'''$$

the $'$ denotes differentiation with respect to $x - ct$. Integrating the above equation yields

$$U'' = -(3U^2 + 4cU + \alpha) = -\frac{\partial V(U)}{\partial U}$$

The expression for $V(U)$ in the above equation is given by

$$V(U) = U^3 + 2cU^2 + \alpha U$$

For constant solutions of U, where $(U' = U'' = 0)$, i.e., the zeros of $\partial V/\partial U$, the discriminant is given by

$$\triangle = 16c^2 - 12\alpha = 4(4c^2 - 3\alpha)$$

For $c^2 > (3/4)\alpha$, there are two constant solutions given by U_a and U_b. These solutions are plotted in Figure 4.6.

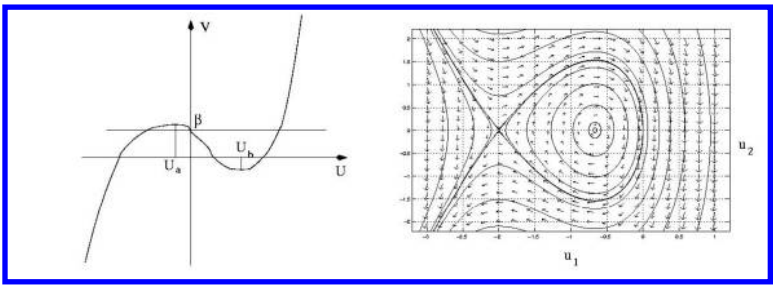


FIGURE 4.6: Simple solution of the *KdV* equation, from Gittelson[1].

4.17.2 Solitary Waves

Here, we discuss a solution of the KdV equation describing solitary waves. The equation is written as

$$\phi_t + \alpha\phi\phi_x + \phi_{xxx} = 0 \tag{4.279}$$

© 2015 by Taylor & Francis Group, LLC

We are interested in looking for steady, progressive wave solutions of the form

$$\phi(x,t) = \phi(\xi), \quad \xi = x - Ut$$

The nonlinear equation reduces to

$$\phi_\xi(\alpha\phi - U) + \phi_{\xi\xi\xi} = 0 \tag{4.280}$$

and

$$|\xi| \to \infty, \quad \phi, \phi_\xi, \phi_{\xi\xi} \to 0$$

Integrating equation (4.280), one gets

$$\phi_{\xi\xi} = U\phi - \frac{\alpha}{2}\phi^2 \tag{4.281}$$

Simple manipulation yields

$$\frac{1}{2}\phi_\xi^2 = \frac{U}{2}\phi^2 - \frac{\alpha}{6}\phi^3 \tag{4.282}$$

The expression for ξ can be written as

$$\sqrt{2}\xi = \int_{\phi_{max}}^{\phi} \frac{d\phi}{\sqrt{U/2\phi^2 - \alpha/6\phi^2}}$$

Introducing simple transformation and performing algebra yield the solution as

$$\phi = \frac{3U}{\alpha}\sec h^2\left[\frac{\sqrt{U}}{2}(x - Ut)\right]$$

The above solution describes a unidirectional solitary wave. Other important properties are

- these waves move with a velocity larger than the phase speed of the gravity waves and
- the width of these waves is inversely proportional to the square root of their amplitude.

4.18 Exercises

1. In a uniformly rotating (2-dim) fluid with angular velocity Ω, write down the relation between the angular velocity and the vorticity.

© 2015 by Taylor & Francis Group, LLC

2. For a steady, 2-dimensional flow of an ideal fluid, subject to a conservative body force g, show that the vorticity ω is constant along a streamline.

3. The enstrophy in two-dimensions is defined as $\epsilon = \int \omega^2 dV$. Show that for an incompressible fluid, the enstrophy is a conserved quantity.

4. The dispersion relation for gravity waves is given by $\omega^2 = gk$. Calculate the velocity of propagation of the wave $U = \partial\omega/\partial k$ and express it in terms of the wavelength.

5. Derive the dispersion relation for gravity waves on an unbounded surface of fluid with depth h.

6. The wave equation for pressure perturbation in a uniformly rotating fluid of angular velocity Ω is given by

$$\frac{\partial^2}{\partial t^2}(\boldsymbol{\nabla}_H)p + 4\Omega^2 \frac{\partial^2 p}{\partial z^2} = 0$$

where the frequency is ω. Derive the equation for the pressure. Discuss the nature of the waves for $\omega < 2\Omega$ and $\omega > 2\Omega$.

7. Consider the dispersion relation of inertial gravity waves given by

$$\omega^2 = \frac{gBk^2 + f^2[m^2 + 1/4H_0^2]}{k^2 + m^2 + 1/4H_0^2}$$

where g is the gravity, B a measure of the static stability, f the measure of rotation, H_0 the scale height, and k and m are wavenumbers

   (a) What happens in the limit of B and $f \to 0$?

   (b) Discuss the dispersion relation for the short wave limit (k is large).

   (c) Discuss the nature of the waves when $B < 0$.

8. For pure gravity waves in a stratified medium, the dispersion relation is given by

$$\omega^2 = \frac{gBk^2}{k^2 + m^2 + 1/4H_0^2}$$

Under what conditions does the frequency ω become $\pm\sqrt{gB}$?

9. For Lamb waves, $\hat{w} = 0$ (vertical component of velocity), the dispersion relation is given by

$$\frac{1}{c_s^2} - \frac{k^2}{\omega^2 - f^2} = 0$$

where c_s is sound speed. Discuss the limiting cases of short and long waves.

© 2015 by Taylor & Francis Group, LLC

10. For Rossby waves, the Coriolis parameter f is written as $f = f_0 + \beta(y - y_0)$ (β plane approximation). If the dispersion relation is

$$\omega = -\frac{\beta k}{k^2 + l^2 + m^2 f_0^2/gB}$$

discuss the condition for the Rossby wave to exist.

11. In the presence of a uniform flow u_0, the frequency is Doppler shifted as

$$\omega = u_0 k - \frac{\beta k}{k^2 + l^2 + m^2 f_0^2/gB}$$

For short Rossby waves, derive the phase and group velocity and discuss their relationship.

© 2015 by Taylor & Francis Group, LLC

Chapter 5

MHD Waves in Uniform Media

5.1 Basic Equations 235
5.2 Sound Waves 238
5.3 Alfven Waves 239
5.4 Shear Alfven Waves 241
5.5 Compressional Alfven Waves 242
5.6 Magneto Acoustic Waves 243
5.7 Internal and Magneto Acoustic Gravity Waves 246
5.7.1 Internal Alfven Gravity Waves 249
5.7.2 Viscous Alfven Gravity Waves 251
5.8 Phase Mixing of Waves 254
5.8.1 Vertical Scale Larger than the Horizontal Scale 258
5.8.2 Uniform Density and Magnetic Field 259
5.9 Resonant Absorption of Waves 260
5.10 Nonlinear Aspects 263
5.11 Exercises 265

5.1 Basic Equations

Waves and oscillations are ubiquitous in nature, when the state of matter is either a liquid or gas. The main restoring forces that act on a gas are (i) gas pressure, (ii) gravity, and (iii) magnetic fields. A gas in a stable equilibrium, which when disturbed will start oscillating. Depending on the external forces, the resulting characteristic oscillations, or modes, are defined as sound, internal gravity, or Alfven waves. The basic physics of the resulting modes are relatively simpler to understand, however, the properties of the oscillations when two or more forces act simultaneously on a gas have a very complicated dependence on the period and wavelength. In the previous chapter, we discussed different types of waves and oscillations, purely from a hydrodynamic point of view. We neglected the effect of magnetic fields in discussing the various modes. However, in this chapter, we shall consider the effect of magnetic fields; the effect of rotation will be ignored for most of the modes discussed in this chapter.

In any theory, the properties of waves are initially determined through the use of first order perturbations to the equations of motion. In MHD (magneto hydrodynamics), we perturb the equations of motion, under the continuum approximation. To begin with, we shall derive the dispersion relation for different types of modes using the standard theory of waves, by linearizing the

© 2015 by Taylor & Francis Group, LLC

basic equations of motion. In this we consider a basic equilibrium situation, perturb it slightly, and assume that the disturbance propagates in the form of a wave. The linearized equations of motion are used along with the perturbed quantities which vary as $\exp[i(\mathbf{k}\cdot\mathbf{r}-\omega t)]$, to derive the dispersion relation, which relates the frequency ω and the wavenumber $\mathbf{k}$.

The basic equations of motion describing the ideal MHD (magnetohydrodynamic) fluid are the continuity of mass, momentum, and energy, together with the induction equation as follows (Priest[154], Satya Narayanan[148]):

$$\frac{D\rho}{Dt}+\rho\nabla\cdot\mathbf{v} = 0 \tag{5.1}$$

$$\rho\frac{D\mathbf{v}}{Dt} = -\nabla p+(\nabla\times\mathbf{B})\times\mathbf{B}/\mu-\rho g\hat{\mathbf{z}}-2\rho\Omega\times\mathbf{v} \tag{5.2}$$

$$\frac{D}{Dt}\left(\frac{p}{\rho^{\gamma}}\right) = 0 \tag{5.3}$$

$$\frac{\partial\mathbf{B}}{\partial t} = \nabla\times(\mathbf{v}\times\mathbf{B}) \tag{5.4}$$

$$\nabla\cdot\mathbf{B} = 0 \tag{5.5}$$

The electric current and temperature have the following forms:

$$\mathbf{j} = \nabla\times\mathbf{B}/\mu \tag{5.6}$$

$$T = \frac{mp}{k_B\rho} \tag{5.7}$$

Here, ρ is the density, $\mathbf{v}$ the velocity, p the pressure, $\mathbf{B}$ magnetic induction, g acceleration due to gravity, Ω the angular velocity, μ magnetic permeability, and γ is the ratio of specific heats.

For simplicity, we shall work in a rotating frame whose angular velocity (Ω) is assumed to be constant relative to an inertial frame. The effect of rotation does not produce significant effect on Maxwell's equations if the absolute speed ($\Omega\times\mathbf{r}+\mathbf{v}$) is less than the velocity of light. Rotation, as we all know gives rise to the Coriolis force given by ($-2\rho\Omega\times\mathbf{v}$) together with a centrifugal force ($1/2\rho\nabla[\Omega\times\mathbf{r}]^2$), not written explicitly in the equations of motion as it can be combined with the gravitational term. The gravitational force $-\rho g\hat{z}$ is assumed to be constant, with the z axis directed along the outward normal to the surface. The ratio p/ρ^{γ} is assumed to be constant following the motions as the plasma is frozen to the magnetic field and thermally isolated from the surroundings.

The equilibrium state will be discussed before linearizing the equations and assuming wavelike solutions. The magnetic field $\mathbf{B}_0$ is assumed to be uniform and permeates a vertically stratified stationary plasma, with a uniform temperature (T_0), and density and pressure distributions as given below:

$$\rho_0(z) = const.\times e^{-z/H} \tag{5.8}$$

$$p_0(z) = const.\times e^{-z/H} \tag{5.9}$$

© 2015 by Taylor & Francis Group, LLC

satisfying the hydrostatic equilibrium (in the absence of any forces and motions):

$$-\frac{dp_0}{dz} - \rho_0 g = 0 \tag{5.10}$$

Here

$$H = \frac{p_0}{\rho_0 g} \tag{5.11}$$

is the scaleheight, which depends on the medium and is not always a constant. The linearized equations of motion are derived under the perturbations of the flow variables:

$$\rho = \rho_0 + \rho_1; \quad \mathbf{v} = \mathbf{v}_1; \quad p = p_0 + p_1; \quad \mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$$

where ρ, $\mathbf{v}$, p, and $\mathbf{B}$ are the density, velocity, pressure, and magnetic field, respectively. The linearized equations of motion (neglecting squares and products of the perturbed quantities) may be written as

$$\frac{\partial \rho_1}{\partial t} + (\mathbf{v}_1 \cdot \nabla)\rho_0 + \rho_0(\nabla \cdot \mathbf{v}_1) = 0 \tag{5.12}$$

$$\frac{\partial p_1}{\partial t} + (\mathbf{v}_1 \cdot \nabla)p_0 - c_s^2\left(\frac{\partial \rho_1}{\partial t} + (\mathbf{v}_1 \cdot \nabla)\rho_0\right) = 0 \tag{5.13}$$

$$\frac{\partial \mathbf{B}_1}{\partial t} = \nabla \times (\mathbf{v}_1 \times \mathbf{B}_0) \tag{5.14}$$

$$\nabla \cdot \mathbf{B}_1 = 0 \tag{5.15}$$

and

$$\rho_0 \frac{\partial \mathbf{v}_1}{\partial t} = -\nabla p_1 + (\nabla \times \mathbf{B}_1) \times \mathbf{B}_0/\mu - \rho_1 g\hat{\mathbf{z}} - 2\rho_0 \Omega \times \mathbf{v}_1 \tag{5.16}$$

where $c_s^2 = \gamma p_0/\rho_0$ is the sound speed. The advantage of linearizing the equations of motion is that by simple algebraic simplifications, the equations of motion involving the perturbed variable can be reduced to an equation in one variable only. Taking the time derivative of equation (5.16) and substituting the terms $\partial\rho_1/\partial t$, $\partial p_1/\partial t$, and $\partial \mathbf{B}_1/\partial t$ from equations (5.12), (5.13), and (5.14), respectively, will result in a single equation as follows:

$$\begin{aligned}\frac{\partial^2 \mathbf{v}_1}{\partial t^2} = {} & c_s^2 \nabla(\nabla \cdot \mathbf{v}_1) - (\gamma - 1) g\hat{\mathbf{z}}(\nabla \cdot \mathbf{v}_1) - g\nabla v_{1z} - 2\Omega \times \frac{\partial \mathbf{v}_1}{\partial t} \\ & + [\nabla \times (\nabla \times (\mathbf{v}_1 \times \mathbf{B}_0)] \times \frac{\mathbf{B_0}}{\mu\rho_0}.\end{aligned} \tag{5.17}$$

For the single equation in $\mathbf{v}_1$, we seek plane-wave solutions of the form

$$\mathbf{v}_1(\mathbf{r}, t) = \mathbf{v}_1 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}$$

Here $\mathbf{k}$ is the wavenumber vector and ω the frequency. The period of the wave can be defined as $2\pi/\omega$ whereas the wavelength λ is just $2\pi/k$. The direction

© 2015 by Taylor & Francis Group, LLC

of propagation of the wave is given by $\mathbf{k}(\equiv \mathbf{k}/k)$. By assuming plane-wave solutions, replace $\partial/\partial t$ by $-i\omega$ and ∇ by $i\mathbf{k}$ in equation (5.17).

Consider the special case when the magnetic field $\mathbf{B}_0 = 0$. In this case, equation (5.17) simplifies to

$$\omega^2 \mathbf{v}_1 = c_s^2 \mathbf{k}(\mathbf{k} \cdot \mathbf{v}_1) + i(\gamma - 1) g \hat{\mathbf{z}}(\mathbf{k} \cdot \mathbf{v}_1) + ig\mathbf{k} v_{1z} - 2i\omega\Omega \times \mathbf{v}_1 \tag{5.18}$$

The dispersion relation $\omega = \omega(\mathbf{k})$, which gives the frequency as a function of the wavenumber $\mathbf{k}$, depends on gravity and magnetic field. Equation (5.18) is a vector equation, with the three velocity components. Once the equation is written in terms of algebraic equations, the dispersion relation can be derived by putting the coefficients of the linearized system and setting the determinant to zero. The velocity $\mathbf{v}_{ph} = (\omega/k)\hat{\mathbf{k}}$ is known as the phase velocity of the wave. Its magnitude (ω/k) gives the speed of propagation in the direction $\hat{\mathbf{k}}$ for a wave specified by a single wavenumber. In most of the problems in wave theory, the dispersion relation among ω, the frequency and wavenumber $\mathbf{k}$ is not necessarily linear. If the relationship is linear, then the waves are termed nondispersive. Otherwise, they are dispersive. For such waves, we can define the concept of a packet (or group) of waves with a range of wavenumbers, traveling with the group velocity $(\mathbf{v}_g)$ with components given by

$$v_{gx} = \frac{\partial \omega}{\partial k_x}; v_{gy} = \frac{\partial \omega}{\partial k_y}; v_{gz} = \frac{\partial \omega}{\partial k_z}$$

The group velocity has the physical interpretation that it is the velocity at which energy is transmitted and, in general, is different in both magnitude and direction from the phase velocity of the wave.

5.2 Sound Waves

The derivation for the phase speed of sound waves may be obtained purely from the hydrodynamic approach also. However, in this section, we shall use the general relation (equation 5.18) and set $g = \mathbf{B_0} = \Omega = 0$, i.e., the effect of gravity, magnetic field, and rotation will be ignored. This will imply that the only restoring force will be the pressure gradient. Equation (5.18), in this case, will reduce to

$$\omega^2 \mathbf{v}_1 = c_s^2 \mathbf{k}(\mathbf{k} \cdot \mathbf{v}_1) \tag{5.19}$$

Taking the scalar product of the above equation with $\mathbf{k}$ and assuming that $\mathbf{k} \cdot \mathbf{v}_1 \neq 0$, we find

$$\omega^2 = k^2 c_s^2 \tag{5.20}$$

The above equation has two solutions. For the propagating disturbances, the dispersion relation for the sound (acoustic) waves is written as

$$\omega = k c_s \tag{5.21}$$

© 2015 by Taylor & Francis Group, LLC

The sound waves propagate equally in all directions (isotropically) with a phase speed given by

$$v_{ph}\left(\equiv \frac{\omega}{k}\right) = c_s \tag{5.22}$$

Differentiating the dispersion relation with respect to k gives the group velocity, which in this case is same as the phase velocity:

$$v_g\left(\equiv \frac{d\omega}{dk}\right) = c_s \tag{5.23}$$

The linear theory of sound waves, in the absence of other restoring forces is rather simple. However, in most cases, compressibility, which is responsible for the sound waves, gets coupled with other forces such as gravity, and magnetic field. Also, when there are no shocks (density and pressure discontinuities), these waves are longitudinal in the sense that the velocity perturbation $\mathbf{v}_1$ is in the direction of the propagation of $\mathbf{k}$, the wavenumber.

5.3 Alfven Waves

It is well known in physics that the tension in an elastic string allows transverse waves to propagate along the string. Similarly, it is reasonable to assume that the magnetic tension will produce transverse waves which will propagate along a magnetic field $\mathbf{B_0}$ with a speed $[(B_0^2/(\mu\rho_0)]^{1/2}$. This expression is termed the Alfven speed given by

$$V_A = \frac{B_0}{(\mu\rho_0)^{1/2}} \tag{5.24}$$

It is well known that the pressure of a gas obeying the adiabatic law, $p/\rho^\gamma = constant$ will produce (longitudinal) sound (acoustic) waves whose phase speed will be $(\gamma p_0/\rho_0)^{1/2}$. In a similar way, one might expect the magnetic pressure $p_m = B_0^2/(2\mu)$ to generate longitudinal magnetic waves propagating across the magnetic field with a phase speed given by $[B_0^2/(\mu\rho_0)]^{1/2}$ which is the Alfven speed .

The Lorentz force $\mathbf{j}\times\mathbf{B}$, on many occasions drives a magnetic wave which can propagate either along or across (it may be oblique) the field. In what follows, we assume that the magnetic field dominates the equilibrium state, so that the restoring forces pressure, gravity and rotation may be neglected. The wave equation in the presence of a magnetic field may be written as

$$\omega^2\mathbf{v}_1 = [\mathbf{k}\times(\mathbf{k}\times(\mathbf{v}_1\times\hat{\mathbf{B}}_0))]\times\hat{\mathbf{B}}_0 V_A^2 \tag{5.25}$$

where $\hat{\mathbf{B}}_0$ is the unit vector in the direction of the magnetic field.

© 2015 by Taylor & Francis Group, LLC

The vector identity on the right hand side of equation(5.25) may be simplified to yield

$$\omega^2 \mathbf{v}_1/V_A^2 = (\mathbf{k}\cdot\hat{\mathbf{B}}_0)^2\mathbf{v}_1 - (\mathbf{k}\cdot\mathbf{v}_1)(\mathbf{k}\cdot\hat{\mathbf{B}}_0)\hat{\mathbf{B}}_0 + [(\mathbf{k}\cdot\mathbf{v}_1) - (\mathbf{k}\cdot\hat{\mathbf{B}}_0)(\hat{\mathbf{B}}_0\cdot\mathbf{v}_1)]\mathbf{k}$$

Assume that the propagation vector of the wave $\mathbf{k}$ makes an angle θ with the equilibrium magnetic field $\mathbf{B}_0$. Then the above equation can be rewritten as

$$\omega^2 \mathbf{v}_1/V_A^2 = k^2\cos^2\theta\mathbf{v}_1 - (\mathbf{k}\cdot\mathbf{v}_1)k\cos\theta\hat{\mathbf{B}}_0 + [(\mathbf{k}\cdot\mathbf{v}_1) - k\cos\theta(\hat{\mathbf{B}}_0\cdot\mathbf{v}_1)]\mathbf{k} \quad (5.26)$$

If we take the scalar product of the above equation with $\hat{\mathbf{B}}_0$, we get

$$\hat{\mathbf{B}}_0\cdot\mathbf{v}_1 = 0 \quad (5.27)$$

In a similar way, if we take the scalar product with respect to $\mathbf{k}$, we get

$$(\omega^2 - k^2V_A^2)(\mathbf{k}\cdot\mathbf{v}_1) = 0 \quad (5.28)$$

which has two distinct solutions.

We shall discuss the simple polarization of Alfven waves. Consider the waves propagating along the z axis. The straight and homogeneous magnetic field will be in the xz plane and has two components:

$$\mathbf{B}_0 = B_0\sin\alpha\hat{x} + B_0\cos\alpha\hat{z}$$

where B_0 is the absolute value of the magnetic field, and α is the angle between the magnetic field and the z axis. In the linear analysis, the set of MHD equations splits into two uncoupled subsets for the different components of the velocity and magnetic field. The first will be the Alfven wave, the other the magneto acoustic waves which will be discussed in detail later in the chapter. The situation where $\mathbf{B}_0 \parallel \hat{\mathbf{z}}$ can lead to two linearized polarized plane Alfven waves, one perturbing V_y, B_y and the other V_x, B_x (Nakariakov [2]).

For perturbations (harmonic) proportional to $\exp(i\omega t - ikz)$, where ω is the frequency and k the wavenumber, a combination of two linearly polarized waves will lead to elliptically polarized Alfven waves given by

$$\begin{aligned} B_y &= A\cos(\omega t - kz) \\ B_x &= B\sin(\omega t - kz) \end{aligned} \quad (5.29)$$

where A and B are constants. The vector of the magnetic field perturbation rotates along an ellipse at the x,y plane. The special case where $A = B$ describes circularly polarized case has $|\mathbf{B}| = constant$. The circularly polarized Alfven waves are an exact solution of the ideal MHD equations for a homogeneous medium. The polarization of Alfven waves is presented in Figure 5.1.

© 2015 by Taylor & Francis Group, LLC

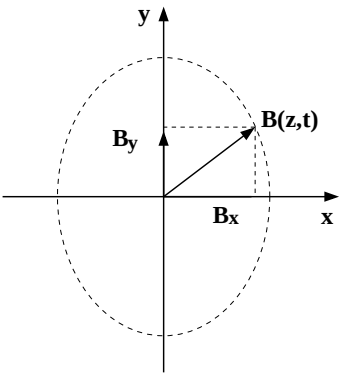


FIGURE 5.1: The magnetic field perturbation in the xy plane, from Nakariakov [2].

5.4 Shear Alfven Waves

For the incompressible case, wherein ($\nabla \cdot \mathbf{v}_1 = 0$), we have

$$\mathbf{k} \cdot \mathbf{v}_1 = 0 \tag{5.30}$$

so that ∇ is replaced by $\mathbf{k}$. Equation (5.26) may be simplified to, after taking the positive square root,

$$\omega = kV_A \cos\theta \tag{5.31}$$

for Alfven waves. This is sometimes referred to as shear Alfven waves. The positive square root describes waves propagating in the same direction as the magnetic field, while the negative root would describe waves propagating in the opposite direction. The variation of the phase speed with θ is presented as a polar diagram which has two circles of diameter V_A (see Figure 5.2).

The group velocity which is obtained by differentiating the dispersion relation with respect to k, gives $\mathbf{v}_g = V_A \hat{\mathbf{B}}_0$. Thus, the energy is transferred at the Alfven speed along the magnetic field, although the individual waves can travel at a different angle of inclination.

In describing the energy of Alfven waves, let us now return to the Lorentz force which is written as

$$\begin{aligned} \mathbf{j}_1 \times \mathbf{B}_0 &= (\mathbf{k} \times \mathbf{B}_1) \times \mathbf{B}_0 \mu \\ &= (\mathbf{k} \cdot \mathbf{B}_0)\mathbf{B}_1/\mu - (\mathbf{B}_0 \cdot \mathbf{B}_1)\mathbf{k}/\mu \end{aligned} \tag{5.32}$$

The first term on the right-hand side of the above equation represents the magnetic tension, while the second term is the magnetic pressure. The driving

© 2015 by Taylor & Francis Group, LLC

force for the Alfven waves is mostly the magnetic tension alone. The ratio of the magnetic energy to the kinetic energy is given by

$$[B_0^2/(2\mu)]/[(1/2)\rho_0 v_1^2]$$

which is almost equal to one. Thus, the Alfven waves contributes to the equipartition between the magnetic and kinetic energy.

5.5 Compressional Alfven Waves

It is simple to realize that one of the solutions of equation (5.28) is

$$\omega = kV_A \tag{5.33}$$

which may be termed compressional Alfven waves. In this case, the phase speed is always V_A for all the angles of propagation. The group velocity is given by $\mathbf{v}_g = V_A\mathbf{k}$, which implies that the energy is propagated isotropically.

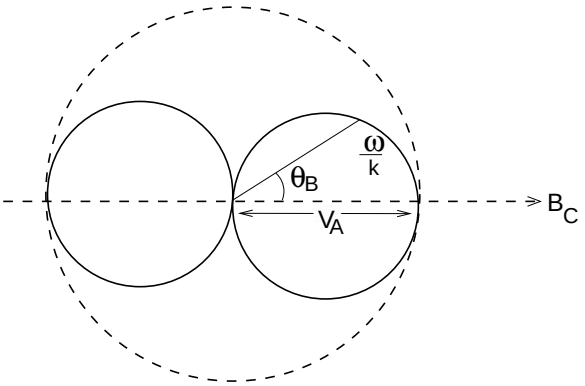


FIGURE 5.2: Polar diagram for the Alfven waves, from Priest [154].

The velocity perturbation $\mathbf{v}_1$ is in the direction normal to $\mathbf{B}_0$ and lies in the plane $(\mathbf{k}, \mathbf{B}_0)$. Thus, it has components along and transverse to $\mathbf{k}$ in general, which will result in changes in both density and pressure. For the special case when $\theta = 0$ (equation 5.31), the compressional Alfven wave is transverse and identical with the ordinary Alfven wave. In this case, the mode is completely dominated by magnetic tension and does not produce any compression. The compressional wave reaches the incompressible limit when the angle of propagation is zero, i.e., along the magnetic field.

© 2015 by Taylor & Francis Group, LLC

5.6 Magneto Acoustic Waves

These waves (as the name suggests) are generated when there is a coupling between the magnetic field and pressure fluctuations. The starting point for a discussion of these waves will be equation (5.18), with the magnetic field included while g and Ω are absent. By simple algebra, one can write down the equation as

$$\frac{\omega^2 \mathbf{v}_1}{V_A^2} = k^2 \cos^2\theta \mathbf{v}_1 - (\mathbf{k}\cdot\mathbf{v}_1)k\cos\theta\hat{\mathbf{B}}_0 + [(1+c_s^2/V_A^2)(\mathbf{k}\cdot\mathbf{v}_1) - k\cos\theta(\hat{\mathbf{B}}_0\cdot\mathbf{v}_1)]\mathbf{k}. \tag{5.34}$$

Taking scalar product of $\mathbf{v}_1$ with $\mathbf{k}$ and $\mathbf{B}_0$, we have

$$(-\omega^2 + k^2c_s^2 + K^2V_A^2)(\mathbf{k}\cdot\mathbf{v}_1) = k^3V_A^2\cos\theta(\hat{\mathbf{B}}_0\cdot\mathbf{v}_1) \tag{5.35}$$

and

$$k\cos\theta c_s^2(\mathbf{k}\cdot\mathbf{v}_1) = \omega^2(\hat{\mathbf{B}}_0\cdot\mathbf{v}_1) \tag{5.36}$$

Eliminating $(\mathbf{k}\cdot\mathbf{v}_1)/(\hat{\mathbf{B}}_0\cdot\mathbf{v}_1)$ between the above two equations leads to the dispersion relation for magneto acoustic (or magneto sonic) waves:

$$\omega^4 - \omega^2k^2(c_s^2+V_A^2) + c_s^2V_A^2k^4\cos^2\theta = 0 \tag{5.37}$$

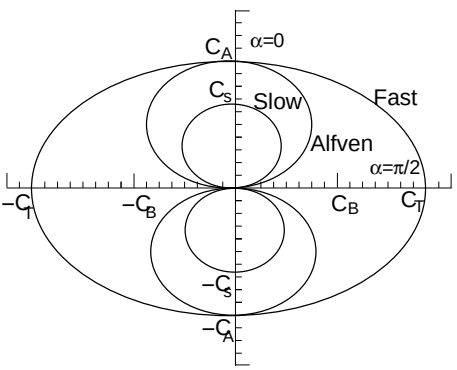


FIGURE 5.3: Polar diagram for the magneto acoustic waves, from Nakariakov [2].

For waves that are outward propagating with $(\omega/k > 0)$, there are two distinct solutions given by

$$\omega/k = \left[(1/2)(c_s^2+V_A^2) \pm (1/2)\sqrt{c_s^4+V_A^4-2c_s^2V_A^2\cos^2\theta}\right]^{1/2} \tag{5.38}$$

© 2015 by Taylor & Francis Group, LLC

One of the frequencies will be higher than the other. The higher-frequency mode is generally termed the fast magneto acoustic wave, while the lower-frequency mode will be called the slow magneto acoustic wave. The phase speed of the Alfven wave lies in between the phase speed of slow and fast modes. Thus, it is referred to in the literature as the intermediate mode.

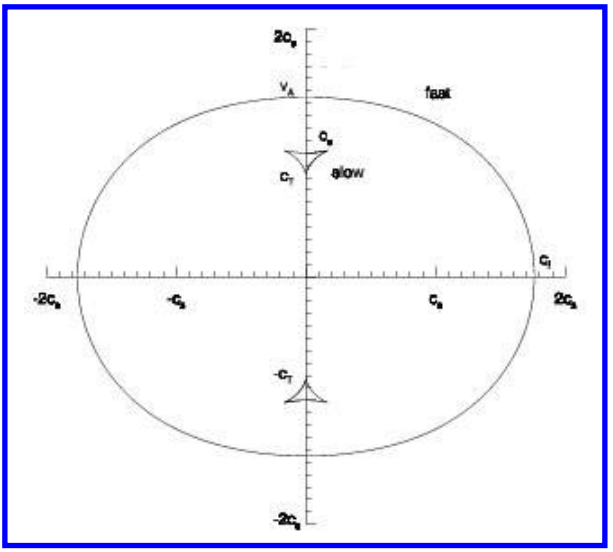


FIGURE 5.4: Group velocity for the magneto acoustic waves, from Nakariakov [2].

It is very clear from the expression of the phase speed of the two magneto acoustic modes that they depend on the direction of propagation. For the case when the propagation is along the magnetic field, the phase speed ω/k is either c_s or V_A, while for propagation across the field ($\theta = \pi/2$), ω/k is $(c_s^2 + V_A^2)^{1/2}$ or 0. As the angle approaches $\pi/2$, the phase speed of the wave tends to

$$c_T = \frac{V_A c_s}{(V_A^2 + c_s^2)^{1/2}}$$

c_T is the component of the phase velocity along the field for propagation almost perpendicular to the field. The wavelength along the field will be larger than the wavelength across the field. c_T is also referred to as the cusp speed for the group velocity of the slow wave. The polar diagram for the magneto acoustic waves and their group velocity are given in Figures 5.3 and 5.4, respectively.

The magneto acoustic wave that we have discussed so far may be considered a sound wave, which gets modified due to the presence of a magnetic field and the compressional Alfven wave, modified by the pressure. The limiting cases of vanishing magnetic field and gas pressure is interesting as in these

© 2015 by Taylor & Francis Group, LLC

cases, the slow mode disappears, while the fast mode becomes the ordinary sound wave and a compressional Alfven wave.

When the plasma $\beta = 2\mu p_0/B_0^2$, the ratio of the magnetic pressure to the gas pressure is larger than unity, then the ratio of the sound speed to the Alfven speed given by c_s^2/V_A^2 is also much larger than one. The dispersion relation for the fast and slow magneto acoustic modes reduces to

$$\omega/k \approx c_s$$

and

$$\omega/k \approx V_A \cos\theta \tag{5.39}$$

respectively. For the slow mode,

$$\mathbf{k} \cdot \mathbf{v} \approx \frac{V_A}{c_s^2} \cos\theta (\mathbf{B}_0 \cdot \mathbf{v}_1)$$

which is much less than unity so that the disturbance is more or less incompressible. The polar diagram for the group velocity of these modes implies that the slow mode energy is seen to propagate in a narrow cone about the magnetic field, while the fast mode has energy which propagates isotropically.

Table 5.1 gives a comparison on the properties of the Alfven and the magneto acoustic (fast and slow) waves. A schematic picture of the dispersion curves for the Alfven, slow, and fast modes is given in Figures (5.5 and 5.6)

TABLE 5.1: MHD Waves

Wave Type	Behavior	Weak Field	Strong Field
Fast	Isotropic $v_{ph} \sim max(v_s, V_A)$	Gas pressure $\mathbf{v} \| \mathbf{k}$	Magnetic pressure $\mathbf{v} \perp \mathbf{B}_0$
Slow	Propagates approximately along $\mathbf{B}_0$ $v_{ph} \sim min(v_s, V_A)$	Magnetic tension $\mathbf{v} \perp \mathbf{k}$	Gas pressure $\mathbf{v} \| \mathbf{B}_0$
Alfven	Propagates along $\mathbf{B}_0$ $v_{ph} = V_A$	Magnetic tension $\mathbf{v} \perp \mathbf{k}$ and $\mathbf{B}_0$	

A schematic picture of the dispersion curves for the Alfven, slow and fast modes is given in Figures (5.5 and 5.6)

© 2015 by Taylor & Francis Group, LLC

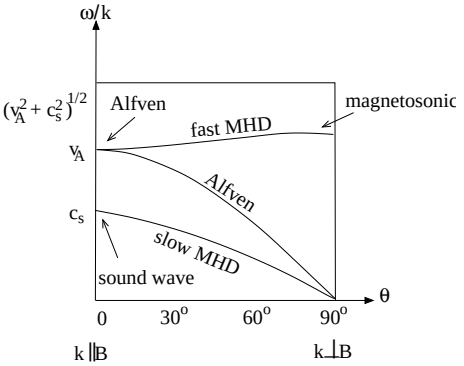


FIGURE 5.5: Dispersion curves for the three modes for the case $V_A < V_s$, from Satya Narayanan [3].

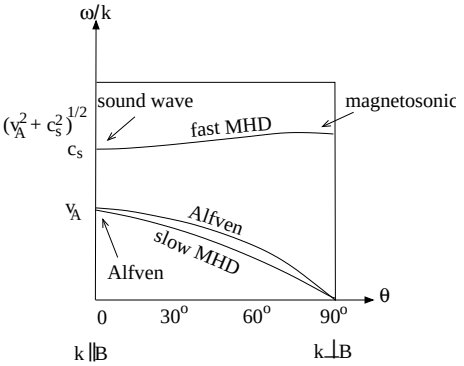


FIGURE 5.6: Dispersion curves for the three modes for the case $V_A > V_s$, from Satya Narayanan [3].

5.7 Internal and Magneto Acoustic Gravity Waves

In studying the properties of internal gravity waves, assume that a parcel of fluid, displaced vertically over a distance from the equilibrium, remains in pressure equilibrium with its surroundings and that the density changes inside the fluid parcel are adiabatic (no heat transfer). The first assumption is valid if the motion is slow compared to the sound waves and the second holds if the motion is fast enough for the entropy to be preserved. The equilibrium is achieved when the pressure gradient is balanced by gravity, i.e.,

$$\frac{dp_0}{dz} = -\rho_0 g \tag{5.40}$$

Outside the fluid parcel, the pressure and density at a height $z + \delta z$ will be

© 2015 by Taylor & Francis Group, LLC

given by $p_0 + \delta p_0$ and $\rho_0 + \delta\rho_0$ so that the above equation may be rewritten as

$$\delta p_0 = -\rho g \delta z \tag{5.41}$$

and

$$\delta\rho_0 = -\frac{d\rho_0}{dz}\delta z \tag{5.42}$$

Since we have assumed that the pressure and density obey the adiabatic relationship, namely, $p/\rho^\gamma = constant$, the change in the pressure will be related to the sound speed in the medium. i.e., $\delta p = c_s^2 \delta\rho$. Substituting the expression for δp into equation (5.41) gives an expression for the internal density change as

$$\delta\rho = -\frac{\rho_0 g \delta z}{c_s^2} \tag{5.43}$$

Performing simple algebra, we can show that

$$g(\delta\rho_0 - \delta\rho) = -N^2 \rho_0 \delta z \tag{5.44}$$

The expression

$$N^2 = -g\Big(\frac{1}{\rho_0}\frac{d\rho_0}{dz} + \frac{g}{c_s^2}\Big) \tag{5.45}$$

when N is real is known as the Brunt–Vaisala frequency (see Chapter 4 also). It plays an important role in studying the static stability of stratified flows. For static stability, $N^2 > 0$.

In the presence of a horizontal magnetic field, the Brunt–Vaisala frequency gets modified to

$$N^2 = -g\Big(\frac{1}{\rho_0}\frac{d\rho_0}{dz} + \frac{g}{c_s^2 + V_A^2}\Big)$$

In the case of uniform temperature (isothermal), N becomes

$$N^2 = \frac{g^2}{c_s^2}\Big(\gamma - \frac{c_s^2}{c_s^2 + V_A^2}\Big)$$

The condition $N^2 > 0$ is usually referred to as the Schwarzschild criterion for convective stability.

In order to derive the dispersion relation for the internal gravity waves, the starting point will be equation (5.18), with the term Ω absent. Taking the scalar product with $\mathbf{k}$ and $\hat{\mathbf{z}}$ in turn and gathering together the terms in v_{1z} and $\mathbf{k}\cdot\mathbf{v}_1$, we obtain

$$\begin{aligned} igk^2 v_{1z} &= (\omega^2 - c_s^2 k^2 + i(\gamma-1)gk_z)(\mathbf{k}\cdot\mathbf{v}_1) \\ (\omega^2 - igk_z)v_{1z} &= (c_s^2 k_z + i(\gamma-1)g)(\mathbf{k}\cdot\mathbf{v}_1) \end{aligned} \tag{5.46}$$

Eliminating $(\mathbf{k}\cdot\mathbf{v}_1)/v_{1z}$ between the above expressions yields

$$(\omega^2 - igk_z)(\omega^2 - c_s^2 k^2 + i(\gamma-)gk_z) = igk^2(c_s^2 k_z + i(\gamma-1)g) \tag{5.47}$$

© 2015 by Taylor & Francis Group, LLC

The frequency of waves we will be interested in should be of the order of the Brunt–Vaisala frequency and much smaller than the frequency of sound waves, which implies that

$$\omega \approx \frac{g}{c_s} \ll kc_s$$

Applying the above condition to equation(5.47), we get

$$\omega^2 c_s^2 \approx (\gamma - 1)g^2(1 - k_z^2/k^2) \tag{5.48}$$

In terms of the inclination angle $\theta = \cos^{-1}(k_z/k)$ and N, the expression can be simplified to yield

$$\omega = N \sin\theta \tag{5.49}$$

The above expression is the relation for internal gravity waves. This is different from surface gravity waves which propagate along the surface between two fluids.

Acoustic and gravity modes occur when both compressibility and buoyancy forces are present together. In general, they remain as distinct modes, but are modified under special circumstances. Taking the scalar product of the basic equation (5.18) with $\mathbf{k}$ and $\hat{\mathbf{z}}$ in turn yields equation (5.46) which in turn gives the dispersion relation:

$$\omega^2(\omega^2 - N_s^2) = (\omega^2 - N^2 \sin^2\theta_g)k'^2 c_s^2 \tag{5.50}$$

where

$$N_s = \frac{\gamma g}{2c_s}\left(\frac{c_s}{2H}\right)$$

$$N = \frac{(\gamma - 1)^{1/2} g}{c_s}$$

$$\sin^2\theta_g = 1 - \frac{k_z^2}{k^2}$$

$$\mathbf{k}' = \mathbf{k} + i\frac{\gamma g}{2c_s^2}\hat{\mathbf{z}}$$

θ_g in equation (5.50) is the angle between the vector $\mathbf{k}'$ and the vertical. $N \geq N_s$, in general, and is equal when $\gamma = 2$.

In the special case when $\omega \ll k'c_s$, the dispersion relation (5.50) reduces to $\omega = N \sin\theta_g$ for gravity waves, whereas for $\omega \gg N$, it becomes $\omega = k'c_s$, pure acoustic waves . For vertical propagation the wave exists only if $\omega > N_s$.

Equation (5.50) can be rewritten as follows:

$$k_z'^2 \omega^2 c_s^2 = \omega^2(\omega^2 - N_s^2) - (\omega^2 - N^2)k_x^2 c_s^2$$

For the case when ω^2 and k_x^2 are positive, there are vertically propagating waves ($k_z'^2 > 0$), provided

$$\omega^2(\omega - n_s^2) > (\omega^2 - N^2)k_x^2 c_s^2$$

© 2015 by Taylor & Francis Group, LLC

The effect of a magnetic field on the acoustic-gravity waves complicates the situation with an extra restoring force and preferred direction in addition to the force due to gravity. The Alfven wave propagates unaltered; however, the magnetic field modifies the propagation characteristics of the acoustic-gravity waves to give rise to two magneto acoustic-gravity waves.

When the Coriolis force is absent, the equation for the wave propagation in a fluid with a uniform magnetic field, temperature and a density proportional to $\exp(-z/H)$ allows the solution

$$\omega^2 = k^2 V_A^2 \cos^2\theta$$

in addition to the two magneto acoustic-gravity modes. In the presence of a magnetic field, plane wave solutions exist when $(kH)^{-1} \ll 1$, i.e., if

$$N_s \ll k c_s$$

which means the wavelength is very much shorter than the scale-height$H$. Using the above expression and equation (5.18), the dispersion relation for the magneto acoustic-gravity waves can be obtained by eliminating $\mathbf{v}_1$ and using the above expression to yield

$$\omega^4 - \omega^2 k^2 (c_s^2 + V_A^2) + k^2 c_s^2 N^2 \sin^2\theta_g + k^4 c_s^2 V_A^2 \cos^2\theta = 0 \tag{5.51}$$

where θ is the angle between $\mathbf{k}$ and the magnetic field. Interesting limiting cases arise. First, when N and V_A are neglected, then the acoustic waves are recovered, while for $V_A = 0$ and $\omega \ll k c_s$, the result for internal gravity is obtained. However, a point of caution is that for vanishing of the Alfven speed, the full dispersion relation for acoustic-gravity waves is not recovered, for the simple reason that there is a coupling between the acoustic and gravity waves.

5.7.1 Internal Alfven Gravity Waves

We shall briefly discuss the derivation of the wave equation for internal-Alfven-gravity waves in stratified shear flows. The wave equation for the motion of a perfectly conducting fluid in the presence of a magnetic field with vertical density stratification will be discussed under the assumption that the motion is two-dimensional, variations being in the x and z directions. The fluid is inviscid, perfectly conducting and adiabatic. The Boussinesq approximation (variation of the vertical coordinate except in the buoyancy term is neglected). The rotation is neglected. The basic shear is assumed to be $(U(z), 0)$. The perturbed velocity components and the magnetic field are (v_x, v_z) (B_x, B_z), respectively, with a uniform magnetic field $(B_0, 0)$, where B_0 is a constant. We also assume that

$$|v_x \partial/\partial x + v_z \partial/\partial z| \ll |\partial/\partial t + U\partial/\partial x| \tag{5.52}$$

$$|B_x \partial/\partial x + B_z \partial/\partial z| \ll |\partial/\partial t + B_0 \partial/\partial x| \tag{5.53}$$

© 2015 by Taylor & Francis Group, LLC

The linearized equations of motion can be written in a matrix form as follows:

$$\begin{pmatrix} \rho_0 D & \rho_0 D_2 U & D_1 & 0 & -\mu B_{x0} D_1 & 0 \\ 0 & \rho_0 D & D_2 & g & 0 & -\mu B_{x0} D_1 \\ D_1 & D_2 & 0 & 0 & 0 & 0 \\ 0 & D_2 \rho & 0 & D & 0 & 0 \\ 0 & 0 & 0 & 0 & D_1 & D_2 \\ -B_{0x} D_1 & 0 & 0 & 0 & D & -D_2 U \\ 0 & -B_{0x} D_1 & 0 & 0 & 0 & D \end{pmatrix} \begin{pmatrix} u \\ w \\ P \\ \rho \\ B_x \\ B_z \end{pmatrix} = 0 \tag{5.54}$$

where P is the perturbed total pressure, that is, the sum of gas and magnetic pressure, ρ is the perturbed density, g is the acceleration due to gravity, $D = \partial/\partial t + UD_1$, $D_1 = \partial/\partial x$ and $D_2 = \partial/\partial z$. By eliminating u, P, ρ, B_x, and B_z from the above matrix equation, we obtain a single wave equation for the vertical component of velocity w as follows:

$$\begin{aligned} &(\partial/\partial t + U\partial/\partial x)^4((v_z)_{xx} + (v_z)_{zz}) - (\partial/\partial t + U\partial/\partial x)^3(U_{zz}(v_z)_x) \\ &+(\partial/\partial t + U\partial/\partial x)^2[N^2(v_z)_{xx} - V_A^2((v_z)_{xxxx} + (v_z)_{xxzz})] \\ &+V_A^2(\partial/\partial t + U\partial/\partial x)[2U_z(v_z)_{xxxz} + U_{zz}(v_z)_{xxx}] - 2V_A^2 U_z^2 (v_z)_{xxxx} = 0 \end{aligned} \tag{5.55}$$

Here, the subscripts x and z denote partial derivatives. Assuming sinusoidal disturbance of the form

$$v_z \exp[i(kx - \omega t)]$$

the above equation reduces to a second order differential equation for the vertical component of the velocity field:

$$\frac{d^2 v_z}{dz^2} - \frac{2k\Omega_A^2 U_z}{\Omega_d(\Omega_d^2 - \Omega_A^2)} \frac{dv_z}{dz} + \left[\frac{k^2 N^2}{(\Omega_d^2 - \Omega_A^2)} + \frac{kU_{zz}}{\Omega_d} - k^2 - \frac{2k^2\Omega^2 A U_z^2}{\Omega_d^2(\Omega_d^2 - \Omega_A^2)} \right] v_z = 0 \tag{5.56}$$

Here, $\Omega_d = \omega - kU$ is the Doppler-shifted frequency, and $\Omega_A = kV_A$ is the Alfven frequency with V_A being the Alfven velocity.

The above equation reduces to the famous Taylor-Goldstein when the magnetic field is neglected, i.e.,

$$\frac{d^2 v_z}{dz^2} + \left[\frac{N^2}{(U-c)^2} - \frac{U_{zz}}{(U-c)} - k^2 \right] v_z = 0 \tag{5.57}$$

U is the basic shear, and $c = \omega/k$ is the phase velocity of the wave. The wave equation for the internal-Alfven-gravity waves in stratified shear flows is singular at $\Omega_d = 0, \pm\Omega_A$, i.e., there are two magnetic singularities in addition to the hydrodynamic singularity. The effects of viscosity as well as thermal and ohmic dissipation may intervene and prevent such singularities.

© 2015 by Taylor & Francis Group, LLC

5.7.2 Viscous Alfven Gravity Waves

In what follows we consider another example of Alfven gravity waves with viscosity. The propagation and dissipation of Alfven gravity waves with viscosity as a damping mechanism will be discussed now. We invoke the WKB and Boussinesq approximations, which will be explained below. The linearized equations of motion can be written as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho_0 u) = 0 \tag{5.58}$$

$$\frac{\partial (\rho_0 u)}{\partial t} = -\nabla (p + \frac{B_0 \cdot B}{\mu_0} - \rho g + \frac{B_0 \cdot \nabla B}{\mu_0} - \nabla \cdot \Pi \tag{5.59}$$

$$\frac{\partial B_\perp}{\partial t} = B_0 \frac{\partial u_\perp}{\partial z} \tag{5.60}$$

$$\frac{\partial B_z}{\partial t} = -B_0 \nabla_\perp \cdot u_\perp \tag{5.61}$$

and

$$\frac{\partial p}{\partial t} + W \frac{dp_0}{dz} = c_s^2 \left(\frac{\partial \rho}{\partial t} + W \frac{d\rho_0}{dz} \right) \tag{5.62}$$

where W is the vertical component of velocity perturbation and $c_s = \sqrt{\gamma p_0 / \rho_0}$ is the sound speed. Taking $q = \rho_0 W$ as the vertical mass flux and N as the Brunt–Vaisala frequency, the above equation can be simplified to yield

$$-\frac{1}{c_s^2} \frac{\partial p}{\partial t} + \frac{\partial \rho}{\partial t} = \frac{N^2}{g} q \tag{5.63}$$

where

$$\frac{N^2}{g} = -\frac{1}{\rho_0} \frac{d\rho_0}{dz} + \frac{1}{\rho_0 c_s^2} \frac{dp_0}{dz} \tag{5.64}$$

If we take $p_m = B_0 B / \mu_0$ for perturbation of the magnetic pressure, then $B_0 \partial / \partial t$ of z-component of induction reduces to

$$\frac{\partial^2 p_m}{\partial t^2} = -\frac{B_0^2}{\mu_0} \frac{\partial}{\partial t} (\nabla_\perp \cdot u_\perp) \tag{5.65}$$

The perpendicular components $\nabla_\perp$ of the momentum equation can be written as

$$\rho_0 \frac{\partial}{\partial t} (\nabla_\perp \cdot u_\perp) = -\nabla_\perp^2 p - \nabla^2 p_m \tag{5.66}$$

Substituting the relation (5.66) into equation (5.65) results in

$$\left(\frac{\partial^2}{\partial t^2} - V_A^2 \nabla^2 \right) p_m = V_A^2 \nabla_\perp^2 P \tag{5.67}$$

V_A is the Alfven velocity. Under the Boussinesq approximation, we ignore the

© 2015 by Taylor & Francis Group, LLC

variation of the density with respect to the spatial coordinate, except when it is coupled with the gravitational term. With this, we have

$$g\frac{\partial \rho}{\partial t} = N^2 q \tag{5.68}$$

and $N^2 = g/H, \quad 1/H = -\frac{1}{\rho_0}\frac{d\rho_0}{dz}$.
The z-component of the momentum equation gives

$$\frac{\partial q}{\partial t} = -\rho g - \frac{\partial p}{\partial z} - \left(\frac{\partial}{\partial x}\pi_x + \frac{\partial}{\partial y}\pi_y + \frac{\partial}{\partial z}\pi_z\right). \tag{5.69}$$

The equation for the total pressure can be written as

$$\begin{aligned}\frac{\partial}{\partial t}\nabla^2(p+p_m) &= -\left[N^2\frac{\partial}{\partial z} + \frac{\partial}{\partial t}\nabla^2\left(\frac{\eta_0 N^2}{3\rho_0 g} + \frac{\eta_0}{\rho_0}\frac{\partial}{\partial z}\right)\right. \\ &\quad -\frac{\partial}{\partial t}\frac{\partial^2}{\partial z^2}\left(\frac{\eta_0 N^2}{3\rho_0 g} + \frac{\eta_0}{\rho_0}\frac{\partial}{\partial z}\right) \\ &\quad \left.+\frac{\partial}{\partial t}\frac{\partial^2}{\partial z^2}\left(-\frac{2\eta_0 N^2}{3\rho_0 g} - \frac{2\eta_0}{\rho_0}\frac{\partial}{\partial z}\right)\right] q \end{aligned} \tag{5.70}$$

Substituting the relation for q into the above equation yields

$$\begin{aligned}&\left(\frac{\partial^2}{\partial t^2} + N^2 - \frac{2\eta_0 N^2}{3\rho_0 g}\frac{\partial}{\partial z}\frac{\partial}{\partial t} - \frac{2\eta_0}{\rho_0}\frac{\partial}{\partial t}\frac{\partial}{\partial z^2}\right)\frac{\partial}{\partial t}\nabla^2\left(\frac{\partial^2}{\partial t^2} - V_A\frac{\partial^2}{\partial z^2}\right)P \\ &= \frac{\partial}{\partial t}\frac{\partial}{\partial z}\left(\frac{\partial^2}{\partial t^2} - V_A^2\nabla^2\right)\left[N^2\frac{\partial}{\partial z} + \frac{\partial}{\partial t}\nabla^2\left(\frac{\eta_0 N^2}{3\rho_0 g} + \frac{\eta_0}{\rho_0}\frac{\partial}{\partial z}\right)\right. \\ &\quad \left.-\frac{\partial}{\partial t}\frac{\partial^2}{\partial z^2}\left(\frac{\eta_0 N^2}{3\rho_0 g} + \frac{\eta_0}{\rho_0}\frac{\partial}{\partial z}\right) + \frac{\partial}{\partial t}\frac{\partial^2}{\partial z^2}\left(-\frac{2\eta_0 N^2}{3\rho_0 g} - \frac{2\eta_0}{\rho_0}\frac{\partial}{\partial z}\right)\right]P \end{aligned} \tag{5.71}$$

For plane waves of the form $\exp i(\omega t - \mathbf{k}\cdot\mathbf{r})$, the dispersion relation can be written as

$$Pk_z^4 + Qk_z^3 + Rk_z^2 + Sk_z + T = 0 \tag{5.72}$$

where

$$P = iV_A^2\omega^2\left(\omega - \frac{i\eta_0 k_\perp^2}{\rho_0}\right), \qquad Q = \frac{i\eta_0 N^2 V_A^2\omega^2 k_\perp^2}{3\rho_0 g}$$

$$R = i\omega^3(V_A^2k_\perp^2 - \omega^2) + \frac{\eta_0\omega^2 k_\perp^2}{\rho_0}(V_A^2k_\perp^2 - 3\omega^2)$$

$$S = \frac{\eta_0 i N^2\omega^2 k_\perp^2}{3\rho_0 g}(V_A^2k_\perp^2 - 3\omega^2), \quad \text{and} \quad T = i\omega^3 k_\perp^2(N^2 - \omega^2)$$

In the absence of viscosity, the above relation reduces to

$$V_A^2k_z^4 + (V_A^2k_\perp^2 - \omega^2)k_z^2 + k_\perp^2(N^2 - \omega^2) = 0 \tag{5.73}$$

© 2015 by Taylor & Francis Group, LLC

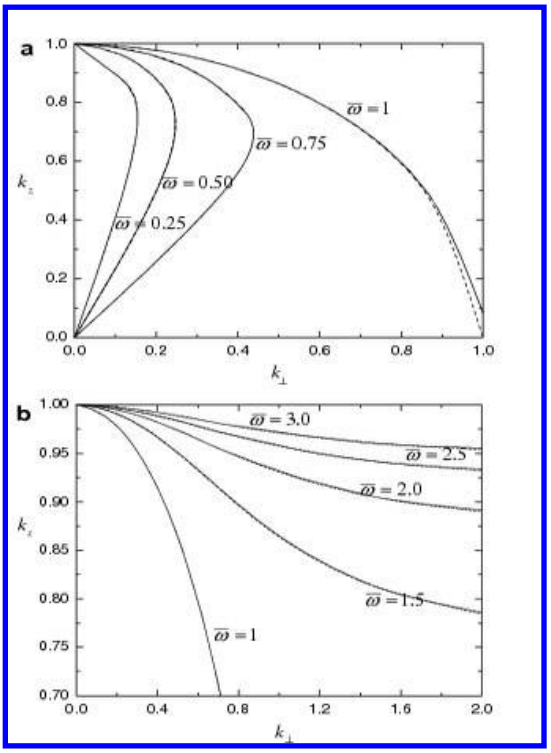


FIGURE 5.7: Wave normal surfaces of hydromagnetic-gravity waves: (a) for wave frequencies less than the Brunt–Vaisala frequency ($\omega < 1$) and (b) for wave frequencies greater than the Brunt–Vaisala frequency ($\omega > 1$), from Pandey and Dwivedi [4].

Here k_z and $k_\perp$ are the wavenumbers in the z-direction and normal to the xz-plane.

When viscosity is included, the wave normal surfaces do not differ significantly from the results obtained in an ideal MHD. From the dissipation point of view, for the wave frequency less than the Brunt–Vaisala frequency (i.e., $\omega < 1$) there are two types of damping: one is strong damping when $L_D \sim \lambda$ (L_D is the damping length) and the other is weak damping when $L_D \gg \lambda$. For the wave frequencies greater than the Brunt–Vaisala frequency (i.e., $\omega > 1$), however, the oscillations are weakly damped. Thus, the Brunt–Vaisala frequency separates regimes of wave frequency in which damping is weak or strong. Figure 5.7 shows the wave normal surfaces at different values of normalized frequencies. From this figure, we infer that in the presence of viscosity (results represented by solid lines), the wave normal surfaces do not differ significantly from the results obtained in an ideal MHD (dashed lines).

Figure 5.8 shows the variation of damping length (represented by solid

© 2015 by Taylor & Francis Group, LLC

lines) and wavelength (dashed lines) as a function of normalized frequency at different values of normalized wave number ($k_\perp$). From this figure we also see that for the lower value of wave number (i.e., $k_\perp = 0.1$ and 0.5), both the wavelength and damping length have a similar nature of variation. For $k_\perp = 1$, however, it has a similar nature only for frequency $\omega > 1.5$ and below this frequency damping length increases while wavelength decreases. We thus find that modes of damping of Alfven gravity waves, whether strong or weak, depend on Brunt–Vaisala frequency and wave number.

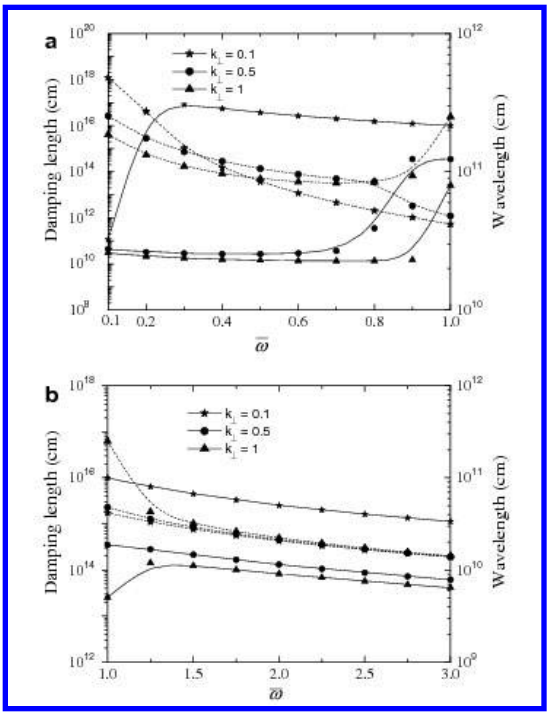


FIGURE 5.8: Variation of damping length (solid lines) and wavelength (dashed lines) as a function of normalized frequency at different values of a normalized wavenumber: (a) for $\omega < 1$ and (b) for $\omega > 1$, from Pandey and Dwivedi[4].

5.8 Phase Mixing of Waves

It is well known (in particular in the photosphere of the Sun) that the magnetic field lines oscillate with a fixed frequency, which leads to the phe-

© 2015 by Taylor & Francis Group, LLC

nomenon of phase mixing. Each field line has its own Alfven speed where the medium is inhomogeneous so that the wave propagates at different phase speeds and in turn may move out of phase. This will result in large spatial gradients until dissipation smoothens them out which will in turn extract energy. With the presence of dissipation, the field line may get heated. This is one mechanism for coronal heating problem. A sketch of the phase mixing is presented in Figure 5.9. An important point is that this heat source will depend on time. In addition, the frequency will vary with spatial distance and the amount of heat deposited.

A variation in the Alfven speed would imply that the waves have different wavelengths on the neighbouring magnetic field lines. The basic equations of motion for viscous MHD equations for an infinitely conducting plasma are given by (Ruderman et al [52]])

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0 \tag{5.74}$$

$$\rho \left[\frac{\partial u}{\partial t} + (u \cdot \nabla) u \right] = -\nabla p + \frac{1}{\mu} (\nabla \times B) \times B \tag{5.75}$$

$$\frac{\partial B}{\partial t} = \nabla \times (u \times B) \tag{5.76}$$

$$\nabla \cdot B = 0 \tag{5.77}$$

Here ρ is the density, p the pressure, u the velocity, and ν the kinematic viscosity of the plasma; B is the magnetic field and μ the magnetic permeability. The gravitational acceleration g is assumed to be constant. In the momentum equation, we consider only shear viscosity and ignore compressional viscosity. The Cartesian coordinate system is adopted with the z axis, antiparallel to the gravitational acceleration. We also assume two-dimensional static equilibrium wherein all quantities depend on x and z only and the y component of the magnetic field is zero.

The equilibrium magnetic field B_0,the plasma pressure p_0, and the density ρ_0 satisfy

$$\nabla p_0 - \frac{1}{\mu} (\nabla \times B_0) \times B_0 + g\rho_0 = 0 \tag{5.78}$$

$$\nabla \cdot B_0 = 0 \tag{5.79}$$

The magnetic field $B_0 = (B_{0x}, 0, B_{0z})$, may be expressed in terms of a magnetic flux function ψ as follows:

$$B_{0x} = B_{00} \frac{\partial \psi}{\partial z}, \quad B_{0z} = B_{00} \frac{\partial \psi}{\partial x}$$

where $B_{00} > 1$ is the field strength at $x = z = 0$. In an open magnetic configuration with $B_{0x} > 0$ everywhere, we have $\partial\psi/\partial x > 0$. The magnetic surfaces are given by the relation $\psi(x, z) = constant$, while the equilibrium pressure and density are related to the flux function as

© 2015 by Taylor & Francis Group, LLC

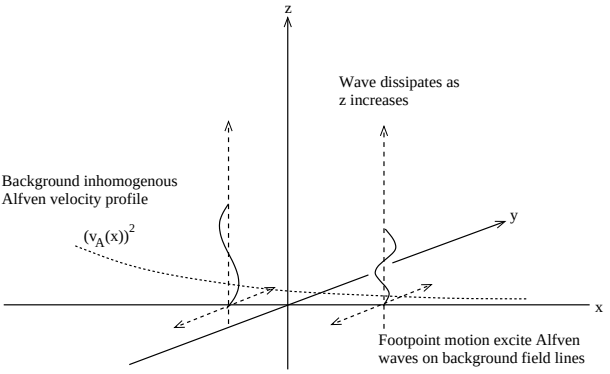


FIGURE 5.9: Phase mixing of Alfven waves, from Ireland and Priest[5].

$$\nabla p_0 = \frac{B_{00}^2}{\mu}\nabla\psi\nabla^2\psi + g\rho_0 \tag{5.80}$$

We assume that the configuration is symmetric with respect to the z axis so that we can write ψ as

$$\psi = \psi_0(Z)x + \epsilon\psi_1(x, Z) \tag{5.81}$$

where $Z = \epsilon z$ and $\epsilon = x_0/H \ll 1$, which is the ratio of the horizontal scale x_0 to the vertical scale H. The x component of equation (5.80) gives the approximate equation

$$\frac{\partial p_0}{\partial x} = \epsilon\frac{B_{00}^2}{\mu}\psi_0(Z)\frac{\partial^2\psi_1}{\partial x^2} \tag{5.82}$$

which can be simplified to yield

$$p_0 = \epsilon\rho_{00}V_{A0}^2\psi_0\frac{\partial\psi_1}{\partial x} + \bar{p_0}(Z) \tag{5.83}$$

where $\bar{p_0}(Z)$ is an arbitrary function, ρ_{00} is the density at $x = z = 0$, and V_{A0} is the Alfven speed at the location. We assume that the plasma beta is of the order of ϵ. The basic equations of MHD are linearized by writing $u = (0, u, 0)$ and $B = B_0 + (0, b, 0)$, which will describe Alfven waves. The y component of the momentum equation and the induction equations yield

$$\rho_0\frac{\partial u}{\partial t} = \frac{1}{\mu}(B_0\cdot\nabla)b + \nabla\cdot(\rho_0\nu\nabla u) \tag{5.84}$$

$$\frac{\partial b}{\partial t} = (B_0\cdot\nabla)u \tag{5.85}$$

Eliminating b from the above equations results in

$$\rho_0\frac{\partial^2 u}{\partial t^2} = \frac{1}{\mu}(B_0\cdot\nabla)^2 + \nabla\cdot\left(\rho_0\nu\nabla\frac{\partial u}{\partial t}\right) \tag{5.86}$$

© 2015 by Taylor & Francis Group, LLC

The above equation may be interpreted as the propagation of Alfven waves in two-dimensional plane magnetic configurations. Let us introduce the function ϕ so that

$$\frac{\partial \psi}{\partial x}\frac{\partial \phi}{\partial x} + \frac{\partial \psi}{\partial z}\frac{\partial \phi}{\partial z} = 0 \tag{5.87}$$

The two functions ψ and ϕ constitute an orthogonal curvilinear coordinate system in the xz plane. The operator $B_0 \cdot \nabla$ can be simplified as

$$B_0 \cdot \nabla = B_{00} J \frac{\partial}{\partial \phi} \tag{5.88}$$

where J is the Jacobian of the coordinate transformation. Using equation (5.88), equation (5.86) can be simplified to yield

$$\sigma \frac{\partial^2 u}{\partial t^2} = V_{A0}^2 J \frac{\partial}{\partial \phi} J \frac{\partial u}{\partial \phi} + J \frac{\partial}{\partial t}\left[\frac{\partial}{\partial \psi}\left(\nu\sigma h_\psi \frac{\partial u}{\partial \psi}\right) + \frac{\partial}{\partial \phi}\left(\nu\sigma h_\psi \frac{\partial u}{\partial \phi}\right)\right] \tag{5.89}$$

where $\sigma = \rho_0/\rho_{00}$ is the dimensionless density. The scale factors h_ψ and h_ϕ are given by

$$h_\psi = J\left[\left(\frac{\partial x}{\partial \phi}\right)^2 + \left(\frac{\partial z}{\partial \phi}\right)^2\right] \tag{5.90}$$

$$h_\phi = J\left[\left(\frac{\partial x}{\partial \psi}\right)^2 + \left(\frac{\partial z}{\partial \psi}\right)^2\right] \tag{5.91}$$

We look for a solution that locally has the form of a propagating wave so that

$$u = u(\theta, \Phi) \tag{5.92}$$

where the phase θ is given by

$$\theta = \omega t - \epsilon^{-1}\Theta(\Phi, \psi) \tag{5.93}$$

Substituting (5.92) and (5.93) into equation (5.89) yields

$$\left[\sigma\omega^2 - V_{A0}^2 J^2 \left(\frac{\partial \Theta}{\partial \Phi}\right)^2\right]\frac{\partial^2 u}{\partial \theta^2} = -\epsilon V_{A0}^2 J\left[\frac{\partial u}{\partial \theta}\frac{\partial}{\partial \Phi}\left(J\frac{\partial \Theta}{\partial \Phi}\right) + 2J\frac{\partial \Theta}{\partial \Phi}\frac{\partial^2 u}{\partial \theta \partial \Phi}\right] + J\epsilon\bar{\nu}\sigma\omega h_\psi \left(\frac{\partial \Theta}{\partial \psi}\right)^2 \frac{\partial^3 u}{\partial \theta^3} + O(\epsilon^2) \tag{5.94}$$

We seek a series solution of the above equation as

$$u = u_1(\theta, \Phi) + \epsilon u_2(\theta, \Phi) + \cdots \tag{5.95}$$

Substituting the above expansion into equation (5.94) gives the first-order approximation

$$\left(\frac{\partial \Theta}{\partial \Phi}\right)^2 = \frac{\omega^2 \sigma}{V_{A0}^2 J^2} \tag{5.96}$$

© 2015 by Taylor & Francis Group, LLC

For upward propagating waves ($\Theta > 0$), we have

$$\Theta = \int_0^{\Phi} \frac{\omega\sigma^{1/2}}{V_{A0}J} d\Phi + \Theta_0(\Phi, \psi) \tag{5.97}$$

Let us assume that $\phi = 0$ at $x = x_0$, $z = 0$, and θ is a constant at $\phi = 0$, so that $\Theta_0 = 0$.

The next order (second) approximation can be written as

$$2J\frac{\partial\Theta}{\partial\Phi}\frac{\partial^2 u_1}{\partial\theta\partial\Phi} + \frac{\partial u_1}{\partial\theta}\frac{\partial}{\partial\Phi}\left(J\frac{\partial\Theta}{\partial\Phi}\right) - \frac{\omega\sigma\bar{\nu}h_\psi}{V_{A0}^2}\left(\frac{\partial\Theta}{\partial\psi}\right)^2\frac{\partial^3 u_1}{\partial\theta^3} = 0 \tag{5.98}$$

Using the relation (5.97), the above equation can be simplified to yield

$$\frac{\partial w}{\partial\Phi} = \lambda\frac{\partial^2 w}{\partial\theta^2} \tag{5.99}$$

where

$$w = \sigma^{1/4}u_1 \tag{5.100}$$

and

$$\lambda(\Phi, \psi) = \frac{\bar{\nu}\sigma^{1/2}h_\psi}{2V_{A0}}(\frac{\partial\Theta}{\partial\psi})^2 \tag{5.101}$$

Equation (5.99) is the diffusion equation in coordinates Φ and θ with the coefficient of diffusion λ depending on Φ and ψ. Consider a monochromatic wave and take w to be proportional to $e^{i\omega t}$. Equation (5.99) reduces to

$$\frac{\partial w}{\partial\Phi} = -\lambda w \tag{5.102}$$

A simple integration of the above equations results in

$$w = W(\psi)\exp(-\int_0^{\phi}\lambda(\Phi', \psi)d\Phi') \tag{5.103}$$

The function $W(\psi)$ can be determined from the boundary condition at $\phi = 0$. The damping is determined by the integral term in the exponent of the above equation.

5.8.1 Vertical Scale Larger than the Horizontal Scale

We consider a magnetic plasma configuration with a vertical characteristic scale larger than the horizontal characteristic scale, The function ψ can by represented by equation (5.81) with the constraint that $|\frac{\partial\psi}{\partial x}| \gg |\frac{\partial\psi}{\partial z}|$. From the orthogonality condition, it follows that $|\frac{\partial\phi}{\partial z}| \gg |\frac{\partial\phi}{\partial x}|$, so that the function ϕ is almost independent of x. Assuming that $B_{0x} \approx B_0$, together with

$$\frac{\partial x}{\partial\phi} = -\frac{1}{J}\frac{\partial\psi}{\partial z}, \qquad \frac{\partial z}{\partial\phi} = \frac{1}{J}\frac{\partial\psi}{\partial x} \tag{5.104}$$

© 2015 by Taylor & Francis Group, LLC

we get the following expressions:

$$J \approx \frac{B_0}{B_{00}} \frac{\partial \phi}{\partial z} \qquad h_\psi \approx \frac{B_0}{B_{00}} \left(\frac{\partial \phi}{\partial z} \right)^{-1} \tag{5.105}$$

With the above expressions, the relation for Θ can be written as

$$\Theta \approx \epsilon \omega \int_0^z \frac{dz'}{V_A(x, z')} \tag{5.106}$$

The expression for λ may be calculated to yield

$$\lambda \approx \epsilon^{-1} \frac{\nu \omega^2 B_0^2}{2 V_A B_{00}^2} \frac{\partial z}{\partial \phi} I^2 \tag{5.107}$$

where

$$I(x, z) = \int_0^z \frac{1}{V_A^2} \frac{\partial V_A}{\partial x} dz' \tag{5.108}$$

Finally, the expression for Λ is calculated as

$$\Lambda = \int_0^\Phi \lambda(\Phi', \psi) d\Phi' \approx \frac{\omega^2}{2 B_{00}^2(x, z)} \int_0^z \frac{\nu(x, z') B_0^2(x, z') I^2(x, z')}{V_A(x, z')} \tag{5.109}$$

The energy flux per unit length in the y direction is

$$S \triangle \psi = e^{-2\Lambda} S_0(x) \triangle x \tag{5.110}$$

where $S_0(x) \triangle x$ is the energy flux at $z = 0$, and $\psi = x$. We give below two specific cases of a magnetic field with the assumption that ν is a constant.

Consider $B_{0x} = 0$, $B_{0z} = B_{00}$, and $\rho(x, z) = \hat{\rho}_0(x) e^{-z/H}$ (uniform magnetic field and exponentially decreasing density). The expression for Λ can be written as

$$\Lambda = \bar{\Lambda}(x)(1 - e^{-z/2H})^3 \tag{5.111}$$

where

$$\bar{\Lambda}(x) = \frac{\nu \omega^2 H^3}{3 V_{A0}^3 \rho_{00}^{3/2} \hat{\rho}_0^{1/2}} \left(\frac{d\hat{\rho}_0}{dx} \right)^2 \tag{5.112}$$

When $z \ll H$, the quantity Λ is proportional to $z^3/6H^3$.

5.8.2 Uniform Density and Magnetic Field

The second example for phase mixing is when the density is uniform in the vertical direction and exponentially diverging magnetic field.

Consider $\rho_0 = \rho_0(x)$, $\psi = H \exp(-z/H) \sin(x/H)$, with the equilibrium magnetic field

$$B_x = B_{00} e^{-z/H} \sin \frac{x}{H}, \quad B_z = B_{00} e^{-z/H} \cos \frac{x}{H} \tag{5.113}$$

© 2015 by Taylor & Francis Group, LLC

By simple algebra we can show that the expression for Λ is

$$\Lambda = \bar{\Lambda}\left(\sin h\frac{z}{H} - \frac{z}{H}\right) \tag{5.114}$$

where

$$\bar{\Lambda}(x) = \frac{\nu\omega^2 H^3}{4V_{A0}^3\rho_{00}^{3/2}\rho_0^{1/2}}\left(\frac{d\rho_0}{dx}\right)^2 \tag{5.115}$$

In this case Λ is proportional to z^3.

5.9 Resonant Absorption of Waves

In the presence of narrow layers with steep variations in the background equilibrium, it is common to represent this background by one or more discontinuities with uniform plasma on either side of the discontinuity. Perturbation of such a medium can support wave propagation along the discontinuity with strongly evanescent behaviour in the direction perpendicular to these discontinuities. In general, it is not easy to find such discontinuities. However, physical quantities such as Alfven speed vary continuously across an interface between different layers of plasmas (see Figure 5.10), an inhomogeneous plasma layer that is impinged by waves at its boundary. Whenever a plasma is nonuniform, a continuous spectrum of Alfven and slow waves may exist in ideal MHD which can lead to resonant absorption.

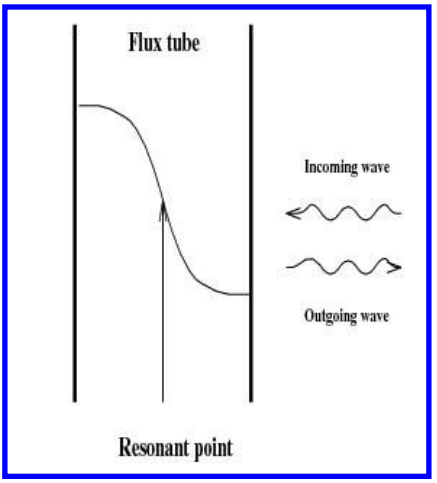


FIGURE 5.10: The inhomogeneous magnetic region with a continuously varying Alfven speed, from Erdelyi[6].

© 2015 by Taylor & Francis Group, LLC

The linearized equations of viscous MHD are used to determine resonant absorption. The magnetic field is assumed to have the form $B(r) = B(r)\hat{e}_z$, which is cylindrically symmetric and the gravity is absent. We also assume that the gas pressure is very much less than the magnetic pressure. The density or Alfven velocity profile is a function of r and z. The equations for radial dependence are given by

$$\rho V_A^2(\omega^2 - \omega_A^2)\frac{d(r\xi_r)}{dr} = -\left[(\omega^2 - \omega_A^2) - \frac{m^2}{r^2}V_A^2\right] rP' \quad (5.116)$$

$$\frac{dP'}{dr} = \rho(\omega^2 - \omega_A^2)\xi_r \quad (5.117)$$

$$\rho(\omega^2 - \omega_A^2)\xi_\phi = \frac{m}{r}P' \quad (5.118)$$

$$\xi_z = 0 \quad (5.119)$$

Here $\bar{\xi}$ is the Lagrangian displacement. P' is the perturbation total pressure, V_A the Alfven speed, ω_A is the Alfven frequency $= k_z V_A$. The eigenvalue problem can be written as

$$\frac{k_i}{\rho_i(\omega^2 - \omega_{Ai}^2)}\frac{I'(k_i R)}{I_m(K_i R)} = \frac{K_e}{\rho_e(\omega^2 - \omega_{Ae}^2)}\frac{K'_m(k_e R)}{K_m(k_e R)} \quad (5.120)$$

Here the indices i and e stand for quantities inside and outside the cylinder. The above relation should be solved for ω for fixed values of m and k_z.

To study resonant absorption, equations (5.116) and (5.117) are the starting point. These equations posses singularity which cannot be avoided. Interestingly, large gradients around the singularity will lead to dissipation. The solutions on the left and right sides of the resonant dissipative layer are connected by the relations

$$||P'|| = 0 \quad (5.121)$$

$$||\xi_r|| = -i\pi\frac{lm^2/r^2}{\omega^2|\rho_i - \rho_e|\alpha}P' \quad (5.122)$$

where $\alpha = \frac{1}{\rho_i - \rho_e}\frac{d\rho}{dr}$(normalized slope of the density profile).

The damping due to resonant absorption is calculated by the dispersion relation at the resonant layer given by

$$D(\omega, k_z, m) = \frac{\xi_{re}}{P'_e} - \frac{\xi_{ri}}{P'_i} = -i\pi\frac{lm^2/r^2}{\omega^2|\rho_i - \rho_e|\alpha} \quad (5.123)$$

Using the Taylor expansion $\omega = \omega_0 + \delta\omega$, we find that

$$\delta\omega = -\frac{\pi l m^2/r_0^2}{\omega_0^2|\rho_i - \rho_e|\alpha\frac{\partial D(\omega_0)}{\partial\omega}} \quad (5.124)$$

The above expression gives the resonant damping of the global mode. Figures

© 2015 by Taylor & Francis Group, LLC

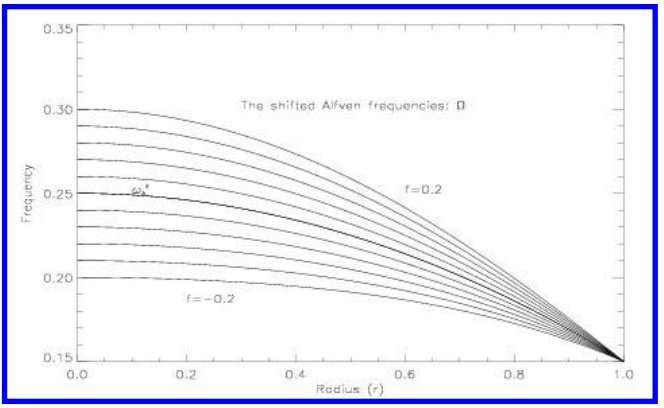


FIGURE 5.11: Doppler shifted Alfven frequency as a function of the radius, from Erdelyi[6].

5.11 and 5.12 describe the Doppler Alfven shifted frequency as a function of the radius from the numerical computations carried out by Erdelyi [6].

The computational results for resonant absorption of Alfven waves for two classes of plasma parameters show that the efficiency of the process strongly depends on the equilibrium parameters and the characteristics of the resonant wave. A steady equilibrium flow can significantly influence the resonant absorption of Alfven waves in magnetic flux tubes. The parametric analysis reveals that the resonant absorption can be strongly enhanced by the equilibrium flow, leading to the total dissipation of the incoming wave.

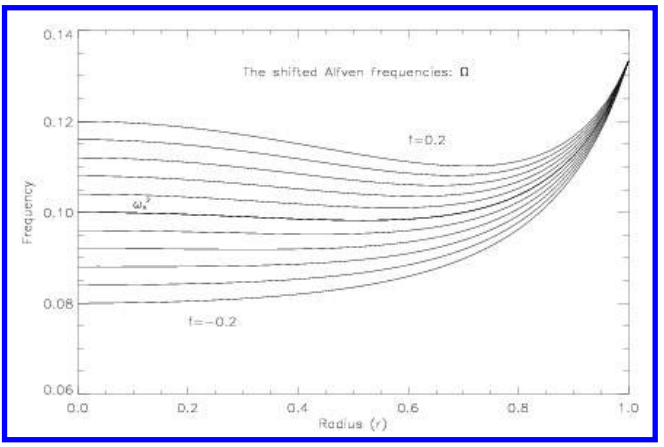


FIGURE 5.12: Same as Figure 5.11 for a different density profile, from Erdelyi[6].

© 2015 by Taylor & Francis Group, LLC

5.10 Nonlinear Aspects

In what follows we briefly discuss the nonlinear behaviour of driven MHD waves in the slow wave dissipative layer. As mentioned above, when the spatial variations of the equilibrium quantities are very rapid in thin regions, the equilibrium state may be represented with uniform plasma on either side of the discontinuity. In such a scenario, new waves modes, called surface waves or surface modes, appear in the plasma. If the plasma is ideal, for some magnetic configurations, there may not be any true normal modes. In such cases, a new phenomenon, called resonant absorption, appears. This is a process which occurs when the global wave motions are locally in resonance with certain magnetic surfaces in the plasma. The advantage with resonance is that it causes energy to build up in the vicinity of the resonant magnetic surfaces, at the expense of the energy in the global motions, which may decay. In such a scenario, the energy density and the spatial gradients become large in the vicinity of the magnetic surfaces. In the presence of dissipation in the system (could be either viscous or electrical resistivity), however small, the transferred energy can be converted into heat. This was one of the mechanisms that has been advocated for heating of the solar coronal plasma.

The theory of resonant absorption has been studied from two points of view. The first point involves a driven problem with an external source of energy which excites plasma oscillations. The other point of view deals with finding solutions of the linearized dissipative MHD equations in the form of surface waves which decay as normal modes.

The nonlinear theory of resonant slow waves in dissipative layers has been studied by Ruderman et al[52]. The detailed analysis is highly technical and in what follows, we sketch the salient points of the theory. Interested readers may look into the original paper for further understanding.

The basic equations of nonlinear visco resistive MHD are given by

$$\frac{\partial \bar{\rho}}{\partial t} + \nabla \cdot (\bar{\rho} u) = 0 \tag{5.125}$$

$$\nabla \cdot B = 0 \tag{5.126}$$

$$\frac{\partial u}{\partial t} + (u \cdot \nabla) u = -\frac{1}{\bar{\rho}} \nabla \bar{P} + \frac{1}{\mu \bar{\rho}} (B \cdot \nabla) B + \nu \nabla^2 u \tag{5.127}$$

$$\frac{\partial B}{\partial t} = (B \cdot u) u - (v \cdot \nabla) B - B \nabla \cdot u + \bar{\eta} \nabla^2 B \tag{5.128}$$

and

$$\frac{\partial}{\partial t} \left(\frac{\bar{p}}{\bar{\rho}^{\gamma}} \right) + u \cdot \nabla \left(\frac{\bar{p}}{\bar{\rho}^{\gamma}} \right) = 0 \tag{5.129}$$

where u is the velocity, B the magnetic induction, $\bar{\rho}$ the density, $\bar{p}$ the pressure, $\bar{P} = \bar{p} + B^2/2\mu$ the total pressure, $\bar{\nu}$ the kinematic coefficient of viscosity, $\bar{\eta}$

© 2015 by Taylor & Francis Group, LLC

the coefficient of magnetic diffusion, μ the magnetic permeability, and γ the adiabatic exponent.

The unperturbed quantities satisfy the condition

$$P_0 = p_0 + \frac{B_0^2}{2\mu} = constant \tag{5.130}$$

We perturb the various quantities as follows:

$$\bar{\rho} = \rho_0 + \rho. \quad \bar{p} = p_0 + p, \quad B = B_0 + b, \quad P = p + \frac{1}{\mu} B_0 \cdot b + \frac{1}{2\mu} b^2$$

In the vicinity of the slow resonant position, the near resonant waves propagate in the z direction with a phase velocity close to $c_T \cos\alpha$. Here c_T is the cusp velocity given by

$$c_t^2 = \frac{c_s^2 V_A^2}{c_s^2 + V_A^2}$$

α is the angle between B_0 and the z axis. Define the total Reynolds number R

$$\frac{1}{R} = \frac{1}{R_e} + \frac{1}{R_m}$$

where R_e is the classical Reynolds number while R_m is the magnetic Reynolds number. We also define a nondimensional quantity ϵ which is a measure of the amplitude far away from the dissipative layer. The method of matched asymptotic expansion (Nayfeh[149]) has been employed wherein the outer and inner expansions are matched in the overlap regions. The solution in the outer region is represented by asymptotic expansions of the form

$$f = \epsilon f^{(1)} + \epsilon^{3/2} f^{(2)} + \cdots \tag{5.131}$$

Substituting the expansion into the system of equations and simplifying leads to a system of two equations for $u^{(1)}$ and $P^{(1)}$ as follows:

$$\frac{\partial u^{(1)}}{\partial x} = \frac{V}{F} \frac{\partial P^{(1)}}{\partial \theta} \tag{5.132}$$

$$\frac{\partial P^{(1)}}{\partial x} = \frac{\rho_0 D_A}{V} \frac{\partial u^{(1)}}{\partial \theta} \tag{5.133}$$

where

$$F = \frac{\rho_0 D_A D_C}{V^4 - V^2(V_A^2 + c_s^2) + V_A^2 c_s^2 \cos^2\alpha}$$

$$D_A = V^2 - V_A^2 \cos^2\alpha$$

$$D_C = (V_A^2 + c_s^2)(V^2 - c_T^2 \cos^2\alpha)$$

The quantities D_A and D_C vanish at Alfven and slow resonant positions,

© 2015 by Taylor & Francis Group, LLC

respectively. The system has two regular singular points. The remaining variables of the first order may be written as

$$\bar{v}_{\perp}^{(1)} = -\frac{V\sin\alpha}{\rho_0 D_A}P^{(1)}, \quad \bar{v}_{||}^{(1)} = \frac{Vc_s^2\cos\alpha}{\rho_0 D_C}P^{(1)} \tag{5.134}$$

$$b_x^{(1)} = -\frac{B_0\cos\alpha}{V}u^{(1)}, \quad b_{\perp}^{(1)} = \frac{B_0\sin\alpha\cos\alpha}{\rho_0 D_A}P^{(1)} \tag{5.135}$$

The single equation for $\bar{v}_{||}^{(1)}$ in terms of $P^{(1)}(\theta)$ with algebraic simplification can be written as

$$\triangle\xi\frac{\partial\bar{v}_{||}^{(1)}}{\partial\theta} - \frac{V[(\gamma+1)V_{Ac}^2+3c_{sc}^2)V_{Ac}^2 cos\alpha}{(V_{Ac}^2+c_{sc}^2]^2}\bar{v}_{||}^{(1)}\frac{\partial\bar{v}_{||}^{(1)}}{\partial\theta}$$

$$+V\left(\nu+\frac{c_{Tc}^2}{V_{Ac}^2}\eta\right)\frac{\partial^2\bar{v}_{||}^{(1)}}{\partial\xi^2} = \frac{Vc_{sc}^2 cos\alpha}{\rho_{0c}(V_{Ac}^2+c_{sc}^2)}\frac{dP^{(1)}}{\partial\theta} \tag{5.136}$$

The subscript denotes the value of the variables at the dissipative layer. In equation (5.136), if $P^{(1)}$ is assumed to be known, then it is a second-order differential equation describing the oscillatory behavior of the parallel flow in the dissipative layer. Once the solution of the above equation is known, then we can determine $u^{(1)}$. If the amplitudes of the perturbations are sufficiently small, we can neglect the nonlinear terms in the above equation. In this case, the solution tends to the linear solution of the ideal viscous MHD equations. The important conclusion of this study is that the absorption of the slow resonant wave in the dissipative layer generates a shear flow parallel to the magnetic surfaces with a characteristic velocity which is of the order of $\epsilon^{1/2}$, where we have defined ϵ previously.

In this chapter, we have talked about waves in a uniform medium (defined as one wherein the flow variables such as velocity, density, pressure, and magnetic field do not have any sudden tangential discontinuities). However, they can have variation with respect to the vertical coordinate, say, z axis, or a radial variation. In the next chapter, we shall deal with a media which permits inhomogeneities in terms of tangential discontinuities, slab and cylindrical geometries embedded in an otherwise uniform medium. It is interesting to note that in addition to the Alfven,and slow and fast magneto acoustic modes, other modes such as the surface, body, kink, sausage, flute modes arise in different geometrical considerations.

5.11 Exercises

1. The linearized magnetic induction equation can be written as

$$\frac{\partial B_1}{\partial t} = \nabla\times(v_1\times B_0)$$

© 2015 by Taylor & Francis Group, LLC

where v_1 and B_1 are the perturbed velocity and magnetic field, respectively. If $\nabla \cdot B_1 = 0$ at time $t = 0$, show that it is true for all time $t > 0$.

2. The dispersion relation for acoustic waves is given by $\omega^2 = c_s^2 k^2$. Derive the phase velocity and group velocity. Discuss the relationship between them.

3. The dispersion relation for MHD waves with resistivity η is given by

$$\omega^2 = \frac{(k \cdot B_0)^2}{\mu \rho_0} - i\eta k^2 \omega$$

Under what conditions, does the above relation describe pure Alfven waves?

4. Consider a loop with magnetic field $B_0 = 12$ Gauss (10^{-3} tesla), $L = 60Mm$ (6×10^7 m), $n = 5 \times 10^{14} m^{-3}$ ($\rho_0 = 8 \times 10^{-13} kgm^{-3}$). Determine i) the Alfven speed, and ii) wavenumber of the fundamental mode. Also calculate the frequency and period of oscillation of the fundamental mode.

5. For the plasma parameters mentioned above, determine the frequency, period, and damping time when the resistivity $\eta = 2ms^{-2}$ is included.

6. The dispersion relation for the magneto acoustic modes is given by

$$\omega^4 - \omega^2 k^2 (c_s^2 + v_A^2) + c_s^2 v_A^2 k^4 \cos^2 \theta = 0$$

Discuss the phase and group velocities of the slow and fast modes. What is the magnitude of the Alfven wave, with respect to the acoustic waves?

7. The dispersion relation for the acoustic gravity mode is given by

$$\omega^2(\omega^2 - N_s^2) = (\omega^2 - N^2 \sin^2 \theta_g) k'^2 c_s^2$$

Under what condition does the above relation reduce to pure gravity modes and acoustic waves?

8. When the acceleration due to gravity g is introduced, how does the relation for magnetoacoustic waves get changed? Describe the limiting cases of no magnetic field and zero gravity.

9. Show that the wave equation for the internal Alfven gravity waves reduces to the famous Taylor−Goldstein equation when the magnetic field is neglected. Briefly discuss the singular points of the wave equation.

10. In the presence of viscosity for Alfven gravity waves, what is the importance of the Brunt–Vaisala frequency?

© 2015 by Taylor & Francis Group, LLC

Chapter 6

MHD Waves in Nonuniform Media

6.1 Waves at a Magnetic Interface 267
6.1.1 Incompressible Medium 270
6.1.2 Compressible Medium 271
6.2 Surface and Interfacial Waves 273
6.2.1 Special Cases 276
6.2.2 Presence of Steady Flows 277
6.3 Tangential Discontinuity with Inclined Fields and Flows 278
6.4 Two-Mode Structure of Alfven Surface Waves 280
6.5 Magneto Acoustic-Gravity Surface Waves with Flows 282
6.6 Waves in a Magnetic Slab 284
6.6.1 Compressible Case 286
6.6.2 Effect of Flows inside the Slab 288
6.6.3 Special Cases 289
6.6.4 Effect of Flows and Gravity with an Application 289
6.7 Negative Energy Waves 292
6.8 Waves in Cylindrical Geometries 293
6.8.1 Different Types of Modes in Cylindrical Geometry 294
6.9 Slender Flux Tube Equations 297
6.10 Waves in Untwisted and Twisted Tubes 299
6.10.1 Oscillations in Annular Magnetic Cylinders 299
6.10.2 Magnetically Twisted Cylindrical Tube 301
6.11 Applications to Coronal Waves 302
6.11.1 Kink Oscillations 302
6.11.2 Sausage Oscillations 305
6.12 Nonlinear Aspects 306
6.13 Exercises 309

6.1 Waves at a Magnetic Interface

In the previous chapter, we discussed the nature and properties of waves in a uniform medium. As mentioned earlier, by a uniform medium, we mean a medium wherein the physical variables such as velocity, density, pressure, and magnetic field do not have discontinuities inside the medium. In this chapter, we shall extend the study of waves to nonuniform media , wherein we allow structures in the magnetic field, discontinuities in density and pressure with finite geometries such as the slab and the cylinder. We will also study the behavior of waves in an annuli of cylinders, twisted magnetic fields. We shall briefly discuss the propagation of nonlinear waves in a magnetically structured

© 2015 by Taylor & Francis Group, LLC

medium. Evolution equations such as the KdV equation, Benjamin−Ono and Burger's equations will also be dealt with briefly.

To begin with, we examine waves in a magnetically structured medium. The governing equations were dealt with in detail in the previous chapter (Chapter 5). We shall briefly discuss the equilibrium state and the linearized equations of motion for completeness.

The equilibrium state may be written as

$$\mathbf{B} = \mathbf{B}_0(x)\hat{z}, \quad \mathbf{p} = \mathbf{p}_0(x), \quad \rho = \rho_0(x), \quad T = T_0(x) \tag{6.1}$$

where the pressure $\mathbf{p}_0$, density ρ_0, and temperature T_0 are embedded in a unidirectional magnetic field B_0.

We shall assume that all the equilibrium quantities are functions of x in the Cartesian system, while the magnetic field is aligned along the z axis. We have from the momentum equation:

$$\frac{d}{dx}\Big[p_0 + \frac{B_0^2}{2\mu}\Big] = 0 \tag{6.2}$$

which means that the total pressure (gas + magnetic) is constant.

Perturbing the equations of mass conservation, momentum, and magnetic induction and the equation of state about the equilibrium leads to

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho_0 \mathbf{v}) = 0 \tag{6.3}$$

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -\nabla\Big[p + \frac{1}{\mu}\mathbf{B}_0 \cdot \mathbf{B}\Big] + \frac{1}{\mu}(\mathbf{B}_0 \cdot \nabla)\mathbf{B} + \frac{1}{\mu}(\mathbf{B} \cdot \nabla)\mathbf{B}_0 \tag{6.4}$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}_0) \tag{6.5}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{6.6}$$

$$\frac{\partial p}{\partial t} + \mathbf{v} \cdot \nabla p_0 = c^2 s\Big[\frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho_0\Big] \tag{6.7}$$

in which the disturbances have smallamplitude (linear) with velocity $\mathbf{v} = (v_x, v_y, v_z)$ and magnetic field $\mathbf{B} = (B_x, B_y, B_z)$, pressure p, and density ρ.

The magnetic pressure is given by $p_m = \frac{1}{\mu}(\mathbf{B}_0 \cdot \mathbf{B})$ and the total pressure $p_T = p + p_m$. The linearized equations of motion may be written as

$$\rho_0\Big[\frac{\partial^2}{\partial t^2} - V_A^2 \frac{\partial}{\partial z^2}\Big]\mathbf{v}_\perp + \mathbf{v}_\perp\Big[\frac{\partial p_T}{\partial t}\Big] = 0 \tag{6.8}$$

and

$$\Big[\frac{\partial^2}{\partial t^2} - c_s^2 \frac{\partial^2}{\partial z^2}\Big]\mathbf{v}_\parallel - c_s^2 \nabla \cdot \Big[\frac{\partial \mathbf{v}_\perp}{\partial z}\Big] = 0 \tag{6.9}$$

Here $\mathbf{v}_\perp = (v_x, v_y, 0)$ and $\mathbf{v}_\parallel = (\mathbf{0}, \mathbf{0}, \mathbf{v_z})$ are the velocities perpendicular and

© 2015 by Taylor & Francis Group, LLC

parallel to the applied magnetic field $\mathbf{B}_0$. The velocity components and the pressure perturbations p_T and p_m are related by

$$\frac{\partial p_T}{\partial t} = -\rho_0\Big[c_s^2\frac{\partial \mathbf{v}_\parallel}{\partial z} + (c_s^2+V_A^2)\nabla\cdot\mathbf{v}_\perp\Big] \tag{6.10}$$

$$\frac{\partial p_m}{\partial t} = \Big[\frac{dp_0}{dx}\Big]v_x - \rho_0 V_A^2\nabla\cdot\mathbf{v}_\perp \tag{6.11}$$

The relations $c_s(x) = (\gamma p_0(x)/\rho_0(x))^{1/2}$ and $V_A(x) = (\mathbf{B}_0(x)/(\mu\rho_0(x)))^{1/2}$ are the sound speed and Alfven speed, respectively.

Introducing sinusoidal (wavelike) disturbance as

$$v_x = v_x(x)e^{i(\omega t - k_y y - k_z z)}$$

the velocity components v_y and v_z can be eliminated and have equations only in terms of $v_x(x)$ and $p_T(x)$ to derive a pair of ordinary differential equations as follows:

$$\frac{dp_T}{dx} + \frac{i\rho_0}{\omega}(\omega^2 - k_z^2V_A^2)v_x = 0 \tag{6.12}$$

$$(\omega^2-k_z^2V_A^2)(\omega^2-k_z^2)\frac{dv_x}{dx}+\frac{i\omega}{\rho_0(c_s^2+V_A^2)}(\omega^2-\omega_S^2(x))(\omega^2-\omega_f^2(x))p_T = 0 \tag{6.13}$$

where

$$\omega_S^2+\omega_f^2 = (k_y^2+k_z^2)(c_s^2+V_A^2)$$

and

$$\omega_S^2\omega_f^2 = k_z^2(k_y^2+k_z^2)c_s^2V_A^2$$

in which ω_S and ω_f are the frequencies corresponding to slow and fast modes, respectively. At this stage, we define the cusp speed c_T as

$$c_T(x) = c_sV_A/(c_s^2+V_A^2)^{1/2}$$

which is subsonic (Mach number less than one) and sub-Alfvenic (Alfvenic Mach number less than one). The pair of ordinary differential equations can be reduced by eliminating either p_T or v_x. Eliminating p_T gives

$$\frac{d}{dx}\left[\frac{\rho_0(c_s^2+V_A^2)(\omega^2-k_z^2V_A^2)(\omega^2-k_z^2c_T^2)}{(\omega^2-\omega_S^2)(\omega^2-\omega_f^2)}\frac{dv_x}{dx}\right]+\rho_0(\omega^2-k_z^2V_A^2)v_x = 0 \tag{6.14}$$

Equation (6.14) may be considered the wave equation for MHD waves in an inhomogeneous medium. A quick observation of the equation makes it clear to us that it has two singularities, an Alfven singularity $\omega^2 = k_z^2V_A^2(x)$ and the cusp singularity $\omega^2 = k_z^2c_T^2(x)$. These singularities may be associated with the occurrence of continuous spectra and may be attributed to the highly anisotropic nature of the Alfven and magneto sonic waves. There are two other expressions which are cut-off points and not singularities, given by $\omega^2 = \omega_S^2(x)$

© 2015 by Taylor & Francis Group, LLC

and $\omega^2 = \omega_f^2(x)$. Corresponding to wave reflection or trapping, these are associated with a change from oscillatory to evanescent behavior in the waves.

Theoretically, one can think of two types of discretely structured media: one possessing a single interface separating two regions with different plasma parameters such as density, pressure, and magnetic field, and the other in which there are two interfaces combining to form a slab or tube of magnetic field which in some sense is different from its environment.

We shall begin with, a discussion on the waves at a magnetic interface for the simple case of an incompressible fluid (the sound velocity approaching ∞). Later, we shall generalize this to include the effects of compressibility, flows, viscosity, nonparallel propagation , and gravity.

6.1.1 Incompressible Medium

In the limit of $c_s \to \infty$ and setting k_y and v_y equal to zero for simplicity, for two dimensional motions $(v_x, 0, v_z)$, the wave equation (6.14) reduces to

$$\frac{d}{dx}\left[\rho_0(k_z^2 V_A^2 - \omega^2)\frac{dv_x}{dx}\right] - k_z^2\rho_0(k_z^2 V_A^2 - \omega^2)v_x = 0 \tag{6.15}$$

We shall divide the magnetic interface into two media, $(x < 0)$ and $(x > 0)$, with the following assumptions :

$$B_0(x) = B_e \quad \text{for} \quad x > 0$$

$$B_0(x) = B_0 \quad \text{for} \quad x < 0$$

A similar assumption will be necessary for the pressure and density. All the values for the pressure, density, and magnetic field are assumed to be uniform with a discontinuity at the interface $x = 0$.

The total pressure balance condition is given by

$$p_0 + \frac{B_0^2}{2\mu} = p_e + \frac{B_e^2}{2\mu} \tag{6.16}$$

where p_e and B_e are the pressure and magnetic field for the medium $x > 0$. Using the ideal gas law, the above relation can be simplified to yield

$$\rho_e(c_e^2 + (1/2)\gamma V_{Ae}^2) = \rho_0(c_0^2 + (1/2)\gamma V_A^2) \tag{6.17}$$

where c_0, c_e, V_A, and V_{Ae} are respectively the sound and Alfven speeds in the media $x < 0$ and $x > 0$. With the above assumptions, the wave equation reduces to

$$\rho_0(k_z^2 V_A^2 - \omega^2)\Big[\frac{d^2 v_x}{dx^2} - k_z^2 v_x\Big] = 0 \tag{6.18}$$

The above equation is valid in both media. We are not interested in the vanishing of $\omega^2 = k_z^2 V_A^2$, but only in the differential operator

$$\frac{d^2 v_x}{dx^2} - k_z^2 v_x = 0 \tag{6.19}$$

© 2015 by Taylor & Francis Group, LLC

The solution of the above equation is easy to obtain. It possesses very simple exponential types of solutions, such as

$$v_x \sim e^{-k_z x}$$

and

$$v_x \sim e^{k_z x}$$

The wave equation being linear, a linear combination of the above solutions will also be a solution. In addition, we impose the condition that v_x is bounded for $x \to \pm\infty$. This will imply that the solution is

$$v_x(x) = \alpha_e e^{-k_z x} \quad \text{for} \quad x > 0$$

$$v_x(x) = \alpha_0 e^{k_z x} \quad \text{for} \quad x < 0 \tag{6.20}$$

satisfying the boundedness of v_x for $|x| \to \infty$. α_0 and α_e, are arbitrary constants.

In order to derive the dispersion relation, we need to impose the boundary conditions at the interface $x = 0$. These two conditions are the continuity of the normal velocity component v_x and the total pressure perturbation at the interface. The pressure continuity condition leads to

$$p_T \sim \rho_0(x)(k_z^2 V_A^2(x) - \omega^2)\frac{dv_x}{dx} \tag{6.21}$$

which will be continuous at $x = 0$. Combining both conditions, the dispersion relation may be simplified to yield

$$\omega^2 = k_z^2 c_k^2 \equiv k_z^2 \left[\frac{\rho_0 V_A^2 + \rho_e V_{Ae}^2}{\rho_0 + \rho_e}\right] \tag{6.22}$$

The dispersion relation (6.22) describes hydromagnetic surface waves. It is evident from the relation that the frequency of these waves depends on the magnitude of the density and the Alfven velocities in both media, implying that a change in the magnetic field or density will result in the propagation of these waves. The wave is confined to the interface with a penetration depth given by k_z^{-1}.

6.1.2 Compressible Medium

In this case, as mentioned in the previous chapter, one would expect the appearance of two additional modes, the magneto sonic modes. Returning to equation (6.14) with the restriction of two-dimensionality $(v_x, 0.v_z)$, we have

$$\frac{d}{dx}\left[\frac{\rho_0(x)(c_s^2(x) + V_A^2(x))(k_z^2 c_T^2(x) - \omega^2)}{(k_z^2 c_s^2(x) - \omega^2)}\frac{dv_x}{dx}\right] - \rho_0(x)(k^2{}_z V_A^2(x) - \omega^2)v_x = 0 \tag{6.23}$$

© 2015 by Taylor & Francis Group, LLC

The relation between p_T and v_x is given by

$$p_T = \frac{i\rho_0(x)}{\omega}(c_s^2(x) + V_A^2(x))\frac{(k_z^2 c_T^2(x) - \omega^2)}{(k_z^2 c_s^2(x) - \omega^2)}\frac{dv_x}{dx} \tag{6.24}$$

In a uniform medium, equation (6.23) reduces to

$$\frac{d^2 v_x}{dx^2} - m_0^2 v_x = 0 \tag{6.25}$$

where

$$m_0^2 = \frac{(k_z^2 c_s^2 - \omega^2)(k_z^2 V_A^2 - \omega^2)}{(c_s^2 + V_A^2)(k_z^2 c_T^2 - \omega^2)} \tag{6.26}$$

For the incompressible case, equation (6.26) reduces to k_z^2.

Let us first consider the simple case of a single interface model. Continuing the analysis, similar to the incompressible case wherein we insist that the mode remains bounded for large values of x, and applying the boundary conditions across $x = 0$, the dispersion relation may be simplified to yield

$$\frac{\omega^2}{k_z^2} = V_A^2 - \Big[\frac{M}{1+M}\Big](V_A^2 - V_{Ae}^2) \tag{6.27}$$

where $M = \rho_e m_0/\rho_0 m_e$. Also, m_e is the value of m_0 in the region $x > 0$. The above dispersion relation describes the behaviour of magneto acoustic surface waves, propagating along the interface $x = 0$. It is important to realize that the dispersion relation mentioned above is transcendental as M is a function of ω, unlike the case of incompressible medium, where the phase speed could be written explicitly in terms of the wavenumbers and the interfacial parameters. Thus, due to compressibility, it may possess more than one solution, for example, the fast and slow magneto acoustic surface waves.

If we assume that one side of the interface is field-free, say, $B_e = 0$, then the above equation reduces to

$$(k_z^2 V_A^2 - \omega^2)m_e = \frac{\rho_e}{\rho_0}\omega^2 m_0 \tag{6.28}$$

where now

$$m_e^2 = k_z^2 - \frac{\omega^2}{c_e^2}$$

The dispersion relation for compressible hydromagnetic waves at a single interface can be simplified to yield

$$\epsilon_1(k,\omega)(m_e^2 + l^2)^{1/2} + \epsilon_2(k,\omega)(m_0^2 + l^2) = 0 \tag{6.29}$$

Here,

$$\epsilon_{1,2}(k,\omega) = k^2 B_{0,e}^2/\mu - \rho_{0,e}\omega^2$$

In order to find the roots of the transcendental equation, which represent

© 2015 by Taylor & Francis Group, LLC

the possible modes of surface wave propagation, it should be noted that the dispersion relation will have real roots only when ϵ_1 and ϵ_2 are of opposite sign, and for exponentially decaying solutions, the terms $M_{0,e}^2 + \tan^2\theta$ should both be positive, where $M_{0,e} = m_{0,e}/k$, and $tan\theta = l/k$. The possible regions for surface wave propagation for some specific values of the interface parameters are presented in Figure 6.1.

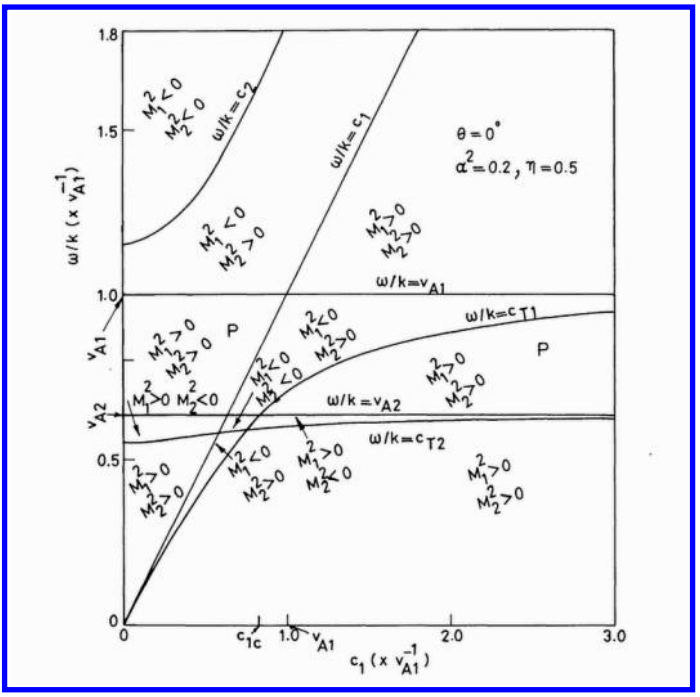


FIGURE 6.1: Possible regions for surface wave propagation for specific values of the interface parameters, from Somasundaram and Uberoi[7].

Considering the phase speed and sound speed plane, ϵ_1 and ϵ_2 are of opposite signs only when

$$min(V_{A1,2}) < \omega/k < max(V_{A1,2})$$

6.2 Surface and Interfacial Waves

In the previous section, the properties of the surface waves when both the wavenumber vector and the magnetic fields on either side of the interface were parallel to the surface $x = 0$ were discussed. In this section, we shall extend

© 2015 by Taylor & Francis Group, LLC

the above results to include the case when the wavenumber and magnetic field vary in both magnitude and direction. We assume that the wavenumber and magnetic field have the following form:

$$\mathbf{k} = (0, k\sin\theta, k\cos\theta)$$

and

$$\mathbf{B}_{01,2} = (0, B_{01,2}\cos\gamma_{1,2}, B_{01,2}\sin\gamma_{1,2})$$

Taking ρ_0 and ρ_e to be the densities, $B_{01,2}$, the magnetic fields (constant) on either side of the interface and c_s and V_A as the sound and Alfven speeds, the dispersion relation by incorporating the boundary conditions at the interface is calculated and written as

$$\tau_1\epsilon_1(\omega,k) + \tau_2\epsilon_2(\omega,k) = 0 \tag{6.30}$$

where

$$\epsilon_1(\omega,k) = \rho_e(-\omega^2 + k^2V_{Ae}^2\sin^2(\theta+\gamma_2)) \tag{6.31}$$
$$\epsilon_2(\omega,k) = \rho_0(-\omega^2 + k^2V_A^2\sin^2(\theta+\gamma_1)) \tag{6.32}$$

and $\tau_{1,2}$ are given by

$$\tau_{1,2}^2 = \frac{\omega^4 - k^2\sin^2(\theta+\gamma_{1,2})(V_{Ae,A}^2 + c_{se,s0}^2) + k^4c_{se,s0}^2V_{Ae,A}^2\sin^4(\theta+\gamma_{1,2})}{k^2c_{se,s0}^2V_{Ae,A}^2\sin^2(\theta+\gamma_{1,2}) - \omega^2(c_{se,s0}^2 + V_{Ae,A}^2)} \tag{6.33}$$

In order to consider the incompressible limit of the above relation, we need to set $c_{se,s0} \to \infty$ and $\tau_{1,2}^2 \to k^2$ so that the dispersion relation reduces to

$$\rho_0(-\omega^2 + k^2V_A^2\sin^2(\theta+\gamma_1)) + \rho_e(-\omega^2 + k^2V_{Ae}^2\sin^2(\theta+\gamma_2)) = 0 \tag{6.34}$$

This gives an analytical expression for the phase speed as

$$\frac{\omega^2}{k^2} = \frac{B_{01}^2\sin^2(\theta+\gamma_1) + B_{02}^2\sin^2(\theta+\gamma_2)}{(\rho_0+\rho_e)} \tag{6.35}$$

We introduce the nondimensional parameters as

$$\eta = \rho_0/\rho_e, \quad \alpha = B_{02}/B_{01}, \quad \omega/kV_A = y, \quad c_{s0}/V_A = x$$

and write $(\theta+\gamma_1) = \phi$, and $(\gamma_1 - \gamma_2) = \chi$,

As already mentioned, the dispersion relation will have real roots only when ϵ_1 and ϵ_2 have opposite signs and $\tau_{1,2}$ should both be positive for the roots to represent surface wave propagation. This implies that for positive roots ω/k should lie in the range

$$min(V_{Ae,A}\sin(\theta+\gamma_{1,2})) < \omega/k < max(V_{Ae,A}\sin(\theta+\gamma_{1,2}))$$

© 2015 by Taylor & Francis Group, LLC

By simplifications, we can show that

$$\epsilon_1 = (-y^2 + \alpha^2\eta\sin^2(\phi-\chi)) = (-y^2 + y_1^2)$$

$$\epsilon_2 = \eta(-y^2 + \sin^2\phi) = (-y^2 + y_2^2)\eta$$

$$\tau_1^2 = \frac{y^4 - y^2(1+x^2) + x^2\sin^2\phi}{x^2\sin^2\phi - y^2(1+x^2)} = \frac{(y^2-y_4^2)(y^2-y_5^2)}{(y_3^2-y^2)}$$

$$\tau_2^2 = \frac{y^4 - y^2(\alpha^2\eta + x^2(c_{s2}^2/c_{s1}^2) + (c_{s2}^2/c_{s1}^2)x^2\alpha^2\sin^2(\phi-\chi)}{(c_{s2}^2/c_{s1}^2)\alpha^2\eta x^2\sin^2(\phi-\chi) - y^2(\alpha^2\eta + x^2(c_{s2}^2/c_{s1}^2))}$$

$$= \frac{(y^2-y_7^2)(y^2-y_8^2)}{(y_6^2-y^2)}$$

Note that y_i are functions of the angles, the sound, and Alfven velocities in both media, respectively.

Let us consider the special case of $c_{s1}/V_A \ll 1$, $B_{01} = 0$, and $\gamma_1 = \gamma_2$. The expressions in the dispersion relation (6.30) reduce to

$$\epsilon_1 = \rho_0(-\omega^2 + k^2V_A^2\sin^2(\theta+\gamma)$$

$$\epsilon_2 = \rho_e(-\omega^2)$$

$$\tau_1^2 = k^2\left(1 - \frac{\omega^2}{k^2V_A^2}\right)$$

$$\tau_2^2 = k^2\left(1 - \frac{\omega^2}{k^2c_{se}^2}\right)$$

Introducing the nondimensional quantities $\alpha = \rho_e/\rho_0$, $\delta = c_{s2}/V_A$,and $x = \omega/kV_A$, and simplifying the relation yields

$$y^3 + Ay^2 + By + C = 0 \tag{6.36}$$

A, B, C are given by

$$A = \frac{\delta^2(\alpha^2-1) - 2\sin^2(\delta+\gamma)}{(1-\alpha^2\delta^2)}$$

$$B = \frac{(1+2\delta^2)\sin^2(\theta+\gamma)}{(1-\alpha^2\delta^2)}$$

$$C = \frac{-\delta^2\sin^4(\theta+\gamma)}{(1-\alpha^2\delta^2)}$$

The condition for the existence of surface waves reduces to

$$min(V_A\sin(\theta+\gamma)) < \omega/k < max(V_A\sin(\theta+\gamma)) \tag{6.37}$$

In what follows, we assume that $\gamma_1 \neq \gamma_2$, with the same expression for wavenumber and magnetic field as mentioned at the beginning of the chapter.

© 2015 by Taylor & Francis Group, LLC

The dispersion relation (6.30) will have coefficients as follows:

$$\epsilon_1(\omega,k) = \rho_0(-\omega^2 + k^2 V_A^2 \sin^2(\theta+\gamma_1)) \quad (6.38)$$

$$\epsilon_2(\omega,k) = \rho_e(-\omega^2 + k^2 V_{Ae}^2 \sin^2(\theta+\gamma_2)) \quad (6.39)$$

and

$$\tau_{1,2}^2 = (\omega^4 - A + B)/(C - D) \quad (6.40)$$

where

$$A = k^2\omega^2(V_{A,Ae}^2 + c_{s0,se}^2)$$

$$B = k^4 c_{s0,se}^2 V_{A,Ae}^2 \sin^2(\theta+\gamma_{1,2})$$

$$C = k^2 c_{s0,se}^2 V_{A,Ae}^2 \sin^2(\theta+\gamma_{1,2})$$

$$D = \omega^2(c_{s0,se}^2 + V_{A,Ae}^2)$$

6.2.1 Special Cases

1. low β plasma with the expressions for $\tau_{1,2}$ reduce to

$$\tau_{1,2}^2 = k^2\left(1 - \frac{\omega^2}{k^2 V_{A,Ae}^2}\right)$$

For the low β case, the pressure balance condition at the interface will yield

$$\rho_0 V_A^2 \approx \rho_e V_{Ae}^2$$

Introducing the nondimensional parameters as mentioned earlier, the relation reduces to

$$(1-\alpha x^2)^{1/2}(\lambda_1^2 - x^2) + (1-x^2)^{1/2}(\lambda_2^2 - \alpha x^2) = 0 \quad (6.41)$$

where $\lambda_{1,2} = \sin(\theta+\gamma_{1,2})$.

2. When $\gamma_1 = \gamma_2 = \pi/2$, the relation simplifies to

$$(1-\alpha x^2)^{1/2}(\cos^2\theta - x^2) + (1-x^2)^{1/2}(\cos^2\theta - \alpha x) = 0 \quad (6.42)$$

which can be further simplified to yield

$$\alpha x^4 - (1+\alpha)x^2 + \cos^2\theta(1+\sin^2\theta) = 0$$

so that

$$x^2 = ((1+\alpha) \pm [(1-\alpha)^2 + 4\alpha\sin^4\theta]^{1/2}/2\alpha$$

© 2015 by Taylor & Francis Group, LLC

3. When $\theta = 0$ this refers to a parallel propagation, and the dispersion relation can be simplified to yield

$$x^6(\alpha^2 - \alpha) + x^4(1 - \alpha^2 + 2\alpha \sin^2 \gamma_1 - 2\alpha \sin^2 \gamma_2)x^2(\sin^4 \gamma_2 - \alpha \sin^4 \gamma_1 + 2\alpha \sin^2 \gamma_2 - 2 \sin^2 \gamma_1) + (\sin^4 \gamma_1 - \sin^4 \gamma_2) = 0 \quad (6.43)$$

4. When $\gamma_1 = \gamma_2 \neq \pi/2$, the solution of the dispersion relation can be written as

$$x^2 = ((1 + \alpha) \pm [(1 - \alpha)^2 + 4\alpha \cos^4 \gamma]^{1/2})/2\alpha$$

5. When the sum of the angles $\theta + \gamma_1$ and $\theta + \gamma_2$ is $\pi/2$, the relation reduces to

$$(1 - \alpha x^2)^{1/2}(1 - x^2) + (1 - x^2)^{1/2}(1 - \alpha x^2) = 0 \quad (6.44)$$

6.2.2 Presence of Steady Flows

The combined effect of nonparallel propagation and steady flow on the properties of hydromagnetic waves will be discussed below. It can in principle give rise to backward propagating surface waves that may be subject to negative energy instabilities. The basic magnetic field is (B_e, B_0) while the steady flow has the form (U_e, U_0) with constant values for magnetic field and velocity shear. The dispersion relation is similar to the one discussed in equation (6.30), with the exception that the frequency gets altered due to the flow. The resulting dispersion relation looks as follows:

$$\rho_e(V_{Ae}^2 - \Omega_e^2)m_0 + \rho_0(V_A^2 - \Omega_0^2)m_e = 0 \quad (6.45)$$

Here,

$$\Omega_e = \omega - kU_e, \qquad \Omega_0 = \omega - kU_0$$

The condition for the existence of surface waves modified by the presence of flow is given by

$$max(v_{ce}^-(\theta), V_A - |U_0|) < c < min(v_{c0}^+(\theta) - |U_0|, V_{Ae})$$

where

$$v_{c(e,0)}^{\pm}(\theta) = \Bigg[(1/2)(V_{A,Ae}^2 + c_{s0,se}^2) \sec^2 \theta \pm (1/2)\Big[(V_{A,Ae}^2 + c_{s0,se}^2)^2 \sec^4 \theta - 4V_{A,Ae}^2 c_{s0,se}^2 \sec^2 \theta\Big]^{1/2}\Bigg]^{1/2} \quad (6.46)$$

Here the + refers to fast waves, the − to slow waves, and $c = \omega/k$.

© 2015 by Taylor & Francis Group, LLC

In the limit of $\theta \to \pi/2$, the phase speed of the surface wave (fast) takes the form

$$c = -\frac{\rho_0}{(\rho_0+\rho_e)}|U_0| + \left[\frac{\rho_0 V^2 A + \rho_e V_{Ae}^2}{(\rho_0+\rho_e)} - \frac{\rho_0\rho_e}{(\rho_0+\rho_e)^2}|U_0^2|\right]^{1/2} \tag{6.47}$$

The phase speed of the forward propagating slow surface wave has the form (in the limit $\theta \to \pi/2$) given by

$$c_a \approx -\Big[\frac{\rho_e}{(\rho_0+\rho_e)}\Big]|U_e| + \left[\frac{\rho_0}{(\rho_0+\rho_e)}V_A^2 - \frac{\rho_0\rho_e}{(\rho_0+\rho_e)^2}|U_e|^2\right]^{1/2} \tag{6.48}$$

6.3 Tangential Discontinuity with Inclined Fields and Flows

The combined effect of nonparallel propagation, steady flow, and inclined magnetic fields on either side of a tangential discontinuity will be examined, with the field strength of the magnetic field being different, though uniform in each layer. The density is also assumed to be different on both sides of the interfacial layer. The interface will in principle support both body waves and surface waves. This will also support fast, Alfven, and slow modes depending on the parametric values of the system.

The equilibrium is given by

$$\frac{d}{dx}\left(p_0 + \frac{B_0^2}{2\mu}\right) = 0$$

The perturbations are as follows :

$$\bar{\rho} = \rho_0(x) + \rho; \quad \bar{\mathbf{v}} = \mathbf{U}(x) + \mathbf{v}; \quad \bar{p} = p_0(x) + P; \quad \bar{\mathbf{B}} = \mathbf{B}(x) + \mathbf{b}$$

where $\mathbf{U} = (0, U_y, U_z)$, and $\mathbf{B} = (0, B_y, B_z)$. The basic equations of MHD can be simplified to get a single differential equation for the velocity component v_x as

$$v_x'' + (m_0^2 + k_y^2)v_x = 0 \tag{6.49}$$

with

$$m_0^2 = \frac{\Omega^2 + k_x^2 c_s^2(\omega_T^2 - \Omega^2)}{c_s^2(\omega_T^2 - \Omega^2)}$$

where Ω, ω, k_x, k_y, and c_s are the Doppler shifted frequency, angular frequency, wavenumbers, and sound speed, respectively.

© 2015 by Taylor & Francis Group, LLC

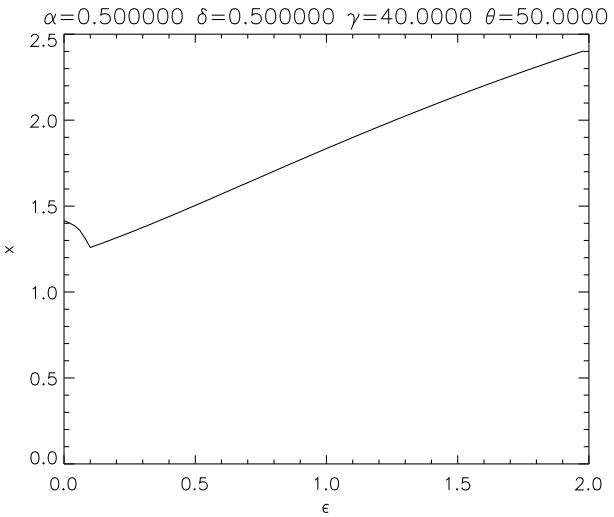


FIGURE 6.2: Dispersive characteristics of surface waves with flows, for specific parametric values, from Satya Narayanan, Pandey, Venkatakrishnan [8].

We shall assume that the variation of the wavenumber is different from that of the magnetic field, i.e.,

$$\mathbf{k} = (0, k\sin\theta, k\cos\theta) \qquad \mathbf{B} = (0, B\sin\gamma, B\cos\gamma)$$

Using similar arguments, the dispersion relation can be shown to be

$$\rho_0[k^2c_{s0}^2\cos^2(\theta-\gamma)-\Omega_0^2](m_e^2+k_y^2)^{1/2}+\rho_e[k^2c_{se}^2\cos^2(\theta-\gamma)-\Omega_e^2](m_0^2+k_y^2)^{1/2}=0 \tag{6.50}$$

Introducing the following nondimensional variables:

$$\alpha=\frac{\rho_e}{\rho_0},\quad \delta=\frac{U_e}{U_0},\quad \epsilon=\frac{U_0}{V_A},\quad x=\frac{\omega}{kV_A}$$

For plasmas with low β, the dispersion relation reduces to

$$\Big[\cos^2(\theta-\gamma)-(x-\epsilon)^2\Big][1-\alpha(x-\epsilon\delta)^2]^{1/2}+\Big[\cos^2(\theta-\gamma)-\alpha(x-\epsilon\delta)^2\Big][1-(x-\epsilon)^2]^{1/2} \tag{6.51}$$

The numerical solution of the dispersion relation for specific values of the plasma parameters is presented in Figures 6.2 and 6.3.

© 2015 by Taylor & Francis Group, LLC

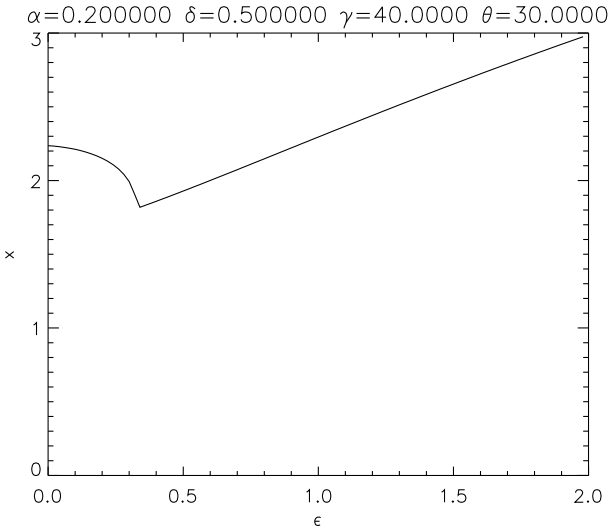


FIGURE 6.3: Same as in the Figure 6.2 for different parametric values.

6.4 Two-Mode Structure of Alfven Surface Waves

In the previous sections, waves and oscillations in structured media were discussed in the absence of viscosity. The Alfven surface waves propagating along a viscous conducting fluid-vacuum interface will be discussed in this section. In addition to the ordinary Alfven surface wave, modified by viscous effects, the interface can support a second mode which is the damped solution of the dispersion equation. The equations of motion for an incompressible viscous fluid of mass density ρ_0, embedded in an external magnetic field $\mathbf{B}_0$, with small perturbations from the equilibrium state can be written as a coupled system of differential equations in v_x and v_z as follows:

$$kD(D^2-\tau^2)v_x + iK^2(D^2-\tau^2)v_z = 0 \tag{6.52}$$

$$k(D^2-\tau^2)v_x + iD(D^2-\tau^2)v_z = 0 \tag{6.53}$$

where D = d/dx, $\tau^2 = K^2 - (i\rho_0/\nu\omega)(\omega^2 - k^2V_A^2)$, and $K^2 = k^2 + l^2$. The above equations can be simplified to yield

$$(D^2-\tau^2)(D^2-K^2)v_x = 0 \tag{6.54}$$

In addition to the boundary conditions we mentioned earlier, namely, the continuity of (1) normal velocity component, and (2) total pressure continuous across the boundary, we must insist on two more boundary conditions, namely, (3) the continuity of the tangential velocity and (4) the tangential viscous

© 2015 by Taylor & Francis Group, LLC

stress. Without loss of generality, we assume that $l = 0$, and the density and viscosity to be negligible at the upper portion of the interface, say, $x > 0$. The dispersion relation can be simplified and written as

$$(x^2-1)^2 - \alpha^2(x^2-1) + i4\nu(x^2-1) + 4\nu^2 T_1 - 4\nu^2 = 0 \tag{6.55}$$

where

$$x = \frac{\omega}{kV_A}, \quad V = \frac{\nu\omega}{\rho_0 V_A^2}, \quad T_1 = k\Big(1 - \frac{i(x^2-1)}{V}\Big)^{1/2}$$

Squaring equation (6.55) and taking the common factor $(x^2 - 1)$ which represents the bulk mode outside, we have

$$x^6 + x^4(-3 - 2\alpha^2 + i8V) + x^2(1+\alpha^2)[(1+\alpha^2) + 2(1-4iV)] - 8V(i+3V)]$$
$$+(1+\alpha^2)[-1-\alpha^2 + 8V(i+V)] + 16V^(1-iV) = 0 \tag{6.56}$$

The effect of uniform flows on the viscous damping of Alfven surface waves at a tangential discontinuity will be a generalization of the above study with the flows included. In this case, the angular frequency gets modified to Doppler-shifted frequency. Introducing the nondimensional parameters as follows:

$$\beta = \frac{B_e}{B_0}, \quad x = \frac{\omega}{kV_A}, \quad R = \frac{U}{V_A}, \quad v = \frac{\nu k}{\rho_0 V_A}$$

which represent the magnetic field ratio, normalized phase velocity, flow velocity, and viscosity, respectively. The relation for the damped Alfven mode (in addition to the surface mode) is written as

$$(x-R)^6 + C(x-R)^5 + D(x-R)^4 + E(x-R)^3 + F(x-R)^2 + G(x-R) + H = 0 \tag{6.57}$$

where

$$C = 6iv$$
$$D = 2ivx - 9v^2 - 2\beta^2 - 3$$
$$E = -14v^2x - 12iv - 6iv\beta^2 - 4v^2R$$
$$F = -v^2x^2 - 2iv\beta^2x - 4ivx - 24iv^3x + 9v^2 + 2(1+\beta^2) + (1+\beta^2)^2$$
$$G = 8iv^3x^3 + 6v^2x + 8v^2x(1+\beta^2) + 12iv^3Rx + 4iv^3R^2 + (6iv + 4v^2R)(1+\beta^2)$$
$$H = v^2x^2 + 2ivx(1+\beta^2) - (1+\beta^2)^2$$

The dispersion relation has to be solved numerically. In the absence of flow, one clearly observes the two-mode structure of Alfven surface waves. However, when the flow is introduced ($R > 0$), the second mode becomes evanescent after a critical value of v and a new mode appears at a higher value of v. The numerical results show that the flow suppresses the existing modes and supports the evolution of new modes. We skip the details of the numerical solutions for the sake of brevity.

© 2015 by Taylor & Francis Group, LLC

6.5 Magneto Acoustic-Gravity Surface Waves with Flows

The effect of gravity on the magneto acoustic surface waves will be discussed in this section. The linear theory of parallel propagation of magneto acoustic-gravity (MAG) surface waves for an interface of a plasma embedded in horizontal magnetic field above a field-free steady plasma medium will be discussed below. The dispersion relation is derived and studied for the case of constant Alfven speed. The presence of new modes called flow or v-modes are observed as a consequence of steady flows. The equilibrium of magneto hydrostatics is given by

$$\frac{d}{dz}\left(p(z) + \frac{B^2(z)}{2\mu}\right) = -g\rho(z) \tag{6.58}$$

In both regions, the assumption of constant Alfven velocity implies

$$\rho(z) = \begin{cases} \rho_0 \exp(-z/H_B), & z > 0 \\ \rho_e \exp(-z/H_e), & z < 0 \end{cases}$$

where H_B and H_e are the density scale-heights in both regions.

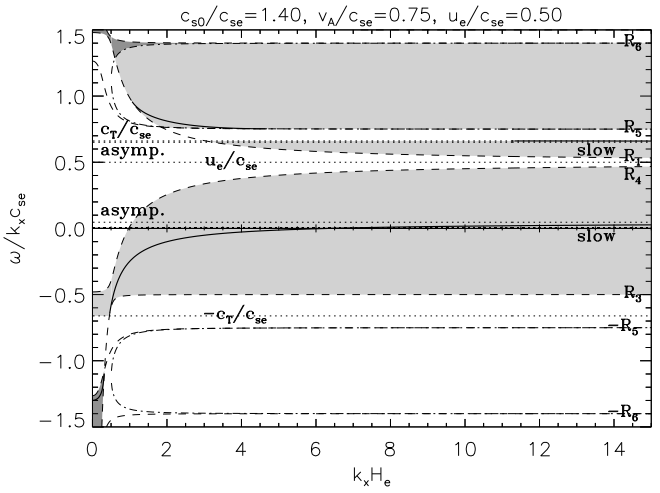


FIGURE 6.4: Solution of the dispersion relation for MAG with flow for specific values, from Erdelyi, Varga and Zetenyi[9].

Two-dimensional linear, isentropic perturbations reduce the linearized compressible ideal MHD equations into a single ordinary differential equation

© 2015 by Taylor & Francis Group, LLC

(ODE) for velocity component v_z as

$$\frac{d}{dz}\left[\frac{\rho(V_A^2+c_s^2)(\Omega^2-k_x^2c_T^2)}{k_x^2c_s^2-\Omega^2}\frac{dv_z}{dz}\right]$$

$$-\left[\rho(\Omega^2-k_x^2V_A^2)+\frac{g^2k_x^2\rho}{k_x^2c_s^2-\Omega^2}+gk_x^2\frac{d}{dz}\left(\frac{\rho c_s^2}{k_x^2c_s^2-\Omega^2}\right)\right]v_z=0 \tag{6.59}$$

Applying the usual boundary conditions, the dispersion relation for MAG can be written as

$$\frac{\omega^2}{k_x^2}=\frac{\rho_0V_A^2}{\rho+\dfrac{\Omega^2}{\omega^2}\rho_e\dfrac{m_0^2(M_e+1/2H_e)}{m_e^2(M_0-1/2H_B)}}-g\frac{\dfrac{\rho_0c_s^2}{k_x^2c_s^2-\omega^2}-\dfrac{\rho_0c_s^2}{k_x^2c_s^2-\Omega^2}}{\dfrac{\rho_0(M_0-1/2H_B)}{m_0^2}+\dfrac{\Omega^2}{\omega^2}\dfrac{\rho(M_e+1/2H_e)}{m_e^2}} \tag{6.60}$$

where

$$m_0^2=\frac{(k_x^2V_A^2-\omega^2)(k_x^2c_s^2-\omega^2)}{(c_s^2+V_A^2)(k_x^2c_T^2-\omega^2)}$$

$$m_e^2=\frac{k_x^2c_s^2-\Omega^2}{c_s^2}$$

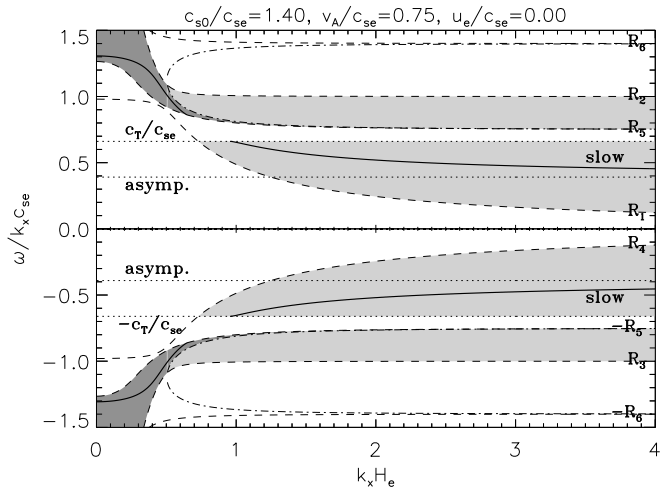


FIGURE 6.5: Solution of the dispersion relation for MAG with flow for specific values, from Varga and Erdelyi[10].

The relation (6.60) describes the parallel propagation of surface waves at a single magnetic interface in a gravitationally stratified atmosphere with the assumption of a constant Alfven speed in the magnetic region and a constant

© 2015 by Taylor & Francis Group, LLC

homogeneous flow in the nonmagnetic region. It is clear from the dispersion relation that the nature of the nondimensionalised phase velocity as a function of the non-dimensional wave number depends on more than on parameter, i.e. the ratio of the sound speeds on either side of the interface, the ratio of the Alfven speed with that of the sound speed (the Alfven Mach number), as well as the nondimensional shear flow, given by u_e/c_{se}. Ω is the Doppler shifted frequency, arising due to a basic flow in the system. More details on the derivation of the dispersion relation is given in Varga and Erdelyi [10].

Solution of the dispersion relation for MAG with flow for some specific values of the parameters are presented in Figures 6.4 and 6.5.

6.6 Waves in a Magnetic Slab

In the previous sections, we discussed waves and oscillations wherein the nonhomogeneity was confined to a single interface (say, a tangential discontinuity). Now, we propose to extend this single discontinuity by a geometry of finite width. In this section, we shall discuss the behavior of waves in a magnetic slab. In a later section, we shall generalize the slab geometry with a cylindrical geometry. This will enable us to study the curvature effects on the waves. To begin with, the equilibrium state is assumed to be

$$\mathbf{B}_0 = \begin{cases} B_0, & |x| < x_0 \\ 0, & |x| > x_0 \end{cases}$$

which is a uniform slab of magnetic field, whose width is $2x_0$ and surrounded by field-free plasma. The wave equation is same as in (6.19) and its solution in different regions can be written as

$$v_x(x) = \begin{cases} \alpha_e e^{-k_z x} & x > x_0 \\ \alpha_0 \cosh(k_z x) + \beta_0 \sinh(k_z x) & |x| < x_0 \\ \beta_e e^{k_z x} & x < -x_0 \end{cases}$$

for arbitrary constants α_e, β_e, α_0, and β_0. The main boundary condition in addition to the boundedness of the solution is that the velocity component and the total pressure are continuous at the boundary $x = \pm x_0$. Figure 6.6 presents different types of modes that are possible in a flux tube. Applying these conditions results in the dispersion relation

$$\frac{k_z^2 V_A^2}{\omega^2} = 1 + \left[\frac{\rho_e}{\rho_0}\right] [\tan h(k_z x_0), \cot h(k_z x_0)] \tag{6.61}$$

Here V_A is the Alfven speed within the slab of gas density ρ_0 surrounded by a field-free medium of gas density ρ_e.

© 2015 by Taylor & Francis Group, LLC

The above relation describes the waves in the slab. Depending on whether the slab is disturbed symmetrically or asymmetrically, one can expect two types of normal modes (see Figure 6.7)

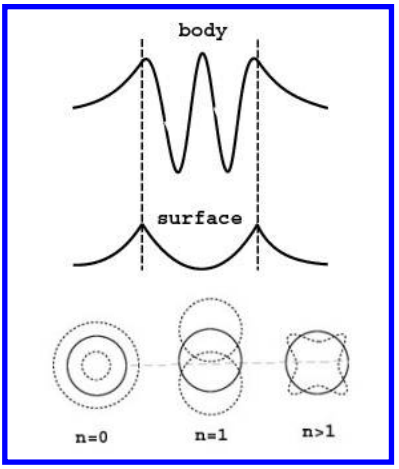


FIGURE 6.6: Different types of modes possible in a flux tube, from Roberts [11].

The hyperbolic function $\tan h$ will represent a symmetric disturbance, which is commonly called a sausage mode. The slab pulsates in such a way that the axis of symmetry remains undisturbed. For the asymmetric mode, the function $\cot h$ will be the corresponding solution. The slab moves back and forth during the wave motion and this mode is usually called the kink mode. The phase velocities of both these modes are less than the Alfven speed of the slab.

It is easy to realize that unlike the hydromagnetic surface wave, the slab waves are dispersive; i.e., the frequency of these modes are functions of the wavenumber and other physical parameters describing the system.

If we assume that the slab is sufficiently long with small radius, it will be called a thin slab. The thin slab approximation corresponds to $k_z x_0 \ll 1$; i.e., the long wavelength disturbances propagate with a phase speed given by

$$\frac{\omega}{k_z} = V_A$$

for the sausage mode and

$$\frac{\omega}{k_z} = \left(\frac{\rho_0}{\rho_e}\right)^{1/2} (k_z x_0)^{1/2}$$

for the kink mode. The opposite situation would be a very wide slab, wherein, $k_z x_0 \gg 1$. In this case, the phase speeds of both the modes coincide and have

© 2015 by Taylor & Francis Group, LLC

the form

$$\frac{\omega}{k_z} = V_A\left[1 + \frac{\rho_e}{\rho_0}\right]^{-1/2}$$

clearly indicating that the behavior of these waves will be similar to an interfacial wave at a single interface which is field free.

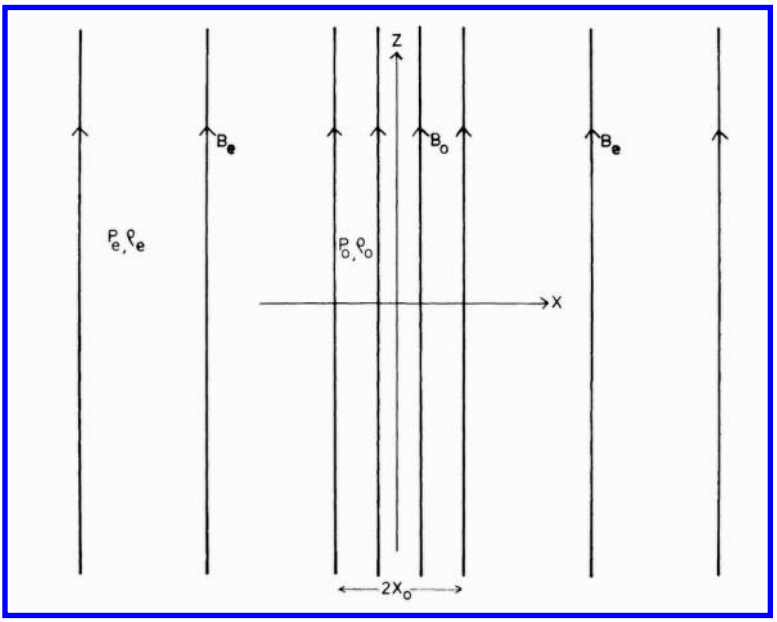


FIGURE 6.7: Magnetic field in a structured slab, from Edwin and Roberts[12].

6.6.1 Compressible Case

The wave equation for the compressible slab geometry is same as the wave equation (6.23). The analysis is similar to that of a single interface. The governing dispersion relation for magneto acoustic slab waves is given by

$$(k_z^2 V_A^2 - \omega)m_e = \left(\frac{\rho_e}{\rho_0}\right)\omega^2 m_0[\tan h, \cot h]m_0 x_0 \tag{6.62}$$

The restriction that m_0^2 should be positive in the case of a single interface, need not be imposed in the slab. However, the condition that $m_e > 0$ still holds as the solution of the above equation must be consistent with the condition $\omega^2 < k_z^2 c_e^2$. Waves with $m_0^2 > 0$ will continue to be called surface waves and those with $m_0^2 < 0$ body waves.

The nomenclature of sausage and kink modes in the incompressible case will continue to hold for the compressible case, corresponding to the tanh and

© 2015 by Taylor & Francis Group, LLC

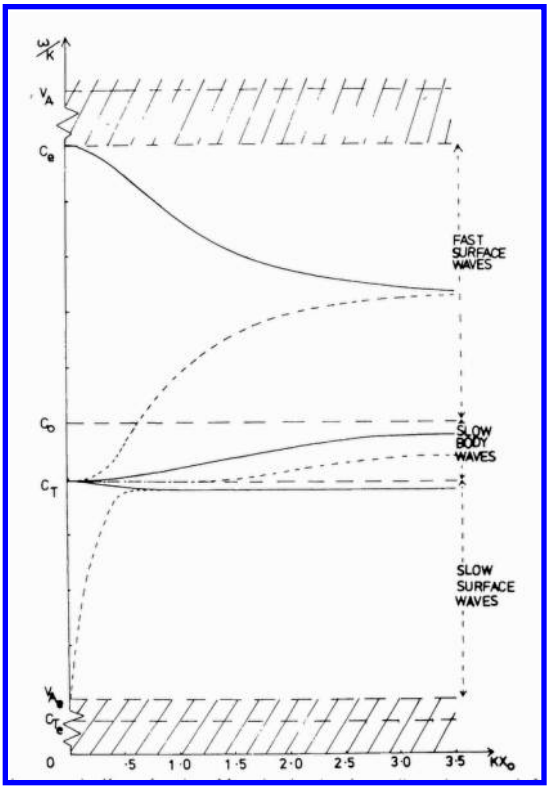


FIGURE 6.8: The phase-speed ω/k_z, plotted as a function of the nondimensional wavenumber $k_z x_0$, from Edwin and Roberts[12].

coth functions, respectively. The solution of the wave equation inside the slab ($x < x_0$) will be of the form

$$v_x = \sinh(m_0 x), \quad \cosh(m_0 x)$$

respectively, for the sausage and kink modes.

The complicated transcendental nature of the dispersion relation prevents us from obtaining the expression for the phase speeds of the modes analytically, and we have to resort to solving the equation numerically. The numerical solution of the dispersion relation is given in Figure 6.8, where the phase speed ω/k_z is plotted as a function of kx_0 for the case when $V_A > c_{se} > c_{s0}$.

In the incompressible limit where the sound speeds tend to infinity, so that m_0 and m_e tend to k_z, the resulting dispersion relation gives rise to two modes, which will be called the sausage surface mode and the kink surface mode. Also in the wide slab limit ($k_z x_0 >> 1$), replacing both the $tanh(m_0 x_0)$ and $coth(m_0 x_0)$ by unity, one recovers the slow surface wave and also a fast surface wave.

© 2015 by Taylor & Francis Group, LLC

The slow surface wave can easily be identified from Figure 6.9 for the thin tube ($k_z x_0 \ll 1$). The slow sausage mode in the long wavelength limit is given by

$$\omega \approx k_z c_T$$

while the kink wave has the same approximate behaviour like the incompressible case.

6.6.2 Effect of Flows inside the Slab

In the above section, we had discussed the behavior of waves in a slab geometry wherein the plasma was at rest, both inside and outside the slab. Now we shall deliberate on the behavior of the waves in a plasma slab which is moving uniformly with respect to the surrounding plasma. We assume that the plasma parameters such as density, pressure, and magnetic field are constant in each layer, but with a discontinuity at the interface separating the slab and the environment. As discussed earlier, the wave equation may be written in terms of either the vertical velocity component or the total pressure. In what follows, we shall discuss the wave equation in terms of the total pressure for an incompressible fluid in a slab moving uniformly relative to the surrounding plasma. The starting point for this discussion will be the wave equation:

$$\nabla^2 \hat{p} = 0 \tag{6.63}$$

where

$$\hat{p} = \bar{p} + \frac{B_{01} \cdot b}{4\pi\mu}$$

Assuming sinusoidal perturbations which have small amplitude, the solution for the wave equation can be written as

$$\hat{p}_1 = A_1 \sinh(kx)$$

where A_1 is an arbitrary constant. The solution for the pressure field for the stationary plasma surrounding the moving plasma is given by

$$\hat{p}_2 = B_1 e^{-kx}$$

where B_1 is an arbitrary constant. Applying the boundary conditions that the total pressure and the normal component of velocity are continuous, the dispersion relation may be simplified to yield an analytical expression for the nondimensional phase velocity as

$$\frac{\omega}{kV_A} = \frac{V \pm ([1 + \eta \tan h(ka)][1 + \beta^2 \tan h(ka)] - \eta V^2 \tan h(ka))^{1/2}}{1 + \eta \tan h(ka)} \tag{6.64}$$

$$\frac{\omega}{kV_A} = \frac{V \pm ([1 + \eta \cot h(ka)][1 + \beta^2 \cot h(ka)] - \eta V^2 \cot h(ka))^{1/2}}{1 + \eta \cot h(ka)} \tag{6.65}$$

© 2015 by Taylor & Francis Group, LLC

where $\beta = B_{02}/B_{01}$ and $\eta = \rho_{02}/\rho_{01}$ are the interface parameters, $V = U/V_A$ is a nondimensional velocity. V_A is the Alfven velocity and a is half the width of the moving plasma column. The first mode (equation 6.64) is the symmetric mode and the second (equation 6.65) is the asymmetric mode.

6.6.3 Special Cases

1. In the limit $ka \to 0$, the above equations (6. 64 and 6.65) with $V = 0$, become
$$\frac{\omega}{kV_A} = 1, \quad \text{and} \quad \frac{\omega}{kV_A} = \sqrt{\beta^2/\eta}$$
for both the symmetric and asymmetric modes, respectively. The symmetric mode is independent of the interface parameters β and η, which is not the case for the asymmetric mode as seen from the above expressions.

2. In the limit $ka \to \infty$, both $\tan h(ka)$ and $\cot h(ka) \to 1$, so that relations (6.64) and (6.65) reduce to
$$\frac{\omega}{kV_A} = \frac{V \pm [(1+\eta)(1+\beta^2) - \eta V^2}{1+\eta} \tag{6.66}$$
The phase velocity of both the modes coincide, unlike the case $ka \to 0$. For $V = 0$, the normalized phase speed reduces to
$$\frac{\omega}{kV_A} = \left[\frac{1+\beta^2}{1+\eta}\right]^{1/2}$$

3. For the case $V = 0$ and $\beta = 0$, no flow and outside magnetic field being absent,
$$\frac{\omega}{kV_A} = \pm\left[\frac{1}{\left[1+\eta(\tan h, \cot h)ka\right]}\right]^{1/2}$$
for the symmetric and asymmetric modes, respectively. If the plasma slab is embedded in vacuum, $\eta = 0$, so that
$$\frac{\omega}{kV_A} = V \pm \left[1+\beta^2(\tan h, \cot h)ka\right]$$

6.6.4 Effect of Flows and Gravity with an Application

It is also well known that gravity waves play an important role in studying the coupling of lower and upper solar atmospheric regions and are therefore of tremendous interdisciplinary interest. The gravity waves in the Sun may

© 2015 by Taylor & Francis Group, LLC

be divided into two types, namely, (i) the internal gravity waves, which are confined to the solar interior, and (ii) the atmospheric gravity waves, which are related to the photosphere and chromosphere, and may be found in the solar corona. In general, the observation of gravity mode oscillations of the Sun would provide a wealth of information about the energy-generating regions, which is poorly understood by probing the p-mode oscillations. The effect of uniform flows and gravity will be discussed here. We have assumed a slab of width '2a', having different densities in each layer and uniform flow in the plasma slab, with the gravity acting downwards. For the sake of brevity, we shall skip the details on the derivation of the dispersion relation. For the sake of simplicity, we have assumed the plasma β to be very small. The relation which governs the MHD waves, coupling the slab and its environment is given by

$$\left[\epsilon_1\epsilon_2 + \epsilon_1\epsilon_g + g\epsilon_1\frac{(\rho_g - \rho_2)}{\omega}\right] +$$
$$\left[\epsilon_1^2 + \epsilon_2\epsilon_g + \epsilon_2 g\frac{(\rho_g - \rho_1)}{\Omega}\epsilon_g g\frac{(\rho_1 - \rho_2)}{\Omega}\right]\tan h(2ka) = 0 \qquad (6.67)$$

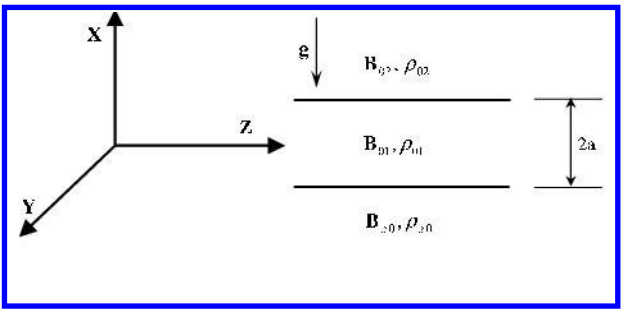


FIGURE 6.9: The geometry, from Satya Narayanan, Kathiravan, and Ramesh[13].

The epsilons and omega are defined as

$$\Omega = \omega - kU_0$$

$$\epsilon_1 = \frac{(\rho_1\Omega)}{k(v_{ph} - \hat{U})^2}[1 - (v_{ph} - \hat{U})^2]$$

$$\epsilon_2 = \frac{(\rho_1\omega)}{kv_{ph}^2}[\beta_1^2 - v_{ph}^2\eta_1]$$

$$\epsilon_g = \frac{(\rho_1 g)}{kv_{ph}^2}[\beta^2 - v_{ph}^2\eta]$$

© 2015 by Taylor & Francis Group, LLC

The following nondimensional quantities are introduced :

$$v_{ph} = \frac{\omega}{k v_{A1}}, \quad \hat{U} = \frac{U_0}{v_{A1}}, \quad \beta = \frac{B_{0g}}{B_{01}}, \quad \beta_1 = \frac{B_{02}}{B_{01}},$$

$$\eta = \frac{\rho_{0g}}{\rho_{01}}, \quad \eta_1 = \frac{\rho_{02}}{\rho_{01}}, \quad G = \frac{g}{k v_{A1}^2}$$

The normalized dispersion relation may be simplified to yield

$$[1 - (v_{ph} - \hat{U})^2][\beta_1^2 - v_{ph}^2 \eta_1) + (\beta^2 - v_{ph}^2 \eta) + G(\eta - \eta_1)]$$

$$\Big[\frac{v_{ph}}{(v_{ph} - \hat{U})}(1 - (v_{ph} - \hat{U})^2)^2 + \frac{(v_{ph} - \hat{U})}{v_{ph}}(\beta_1^2 - v_{ph}^2 \eta_1)(\beta^2 - v_{ph}^2)$$

$$+(\beta_1^2 - v_{ph}^2 \eta_1) G(\eta - 1) + (\beta^2 - v_{ph}^2 \eta) G(1 - \eta_1)\Big] \tan h(2ka) = 0 \qquad (6.68)$$

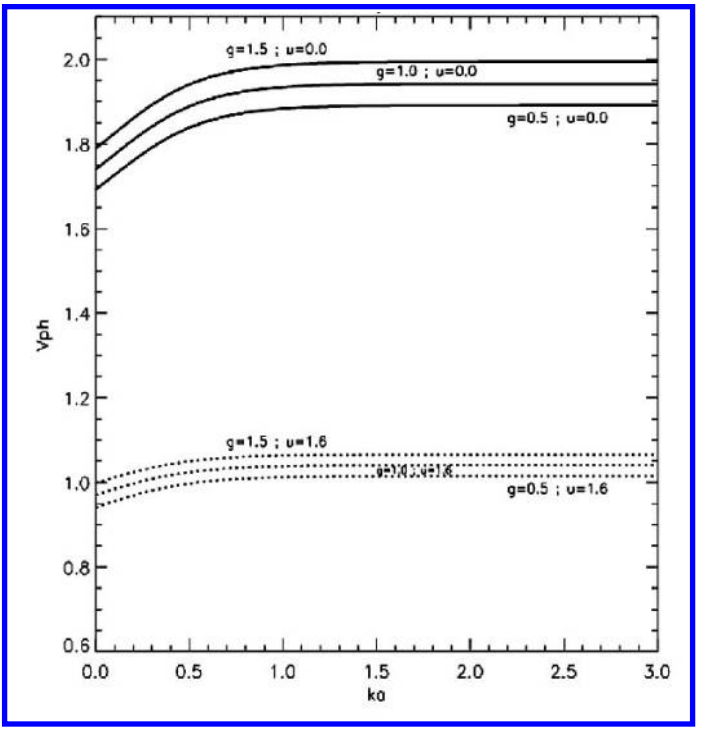


FIGURE 6.10: The normalized phase velocity as a function of the nondimensional wavenumber, from Satya Narayanan, Kathiravan, and Ramesh [13].

The dispersion relation which is solved numerically for the phase speed is plotted as function of the dimensionless wavenumber for various values of the

© 2015 by Taylor & Francis Group, LLC

interface parameters $\beta^2 = 0.5$, $\beta_1^2 = 1.5$, $\eta = 1.8$, $\eta_1 = 1.2$ and different values of the nondimensionalized G and $\hat{U}$ as shown in Figure 6.10. For the fully compressible case, both the fast Alfven gravity surface wave and the slow Alfven gravity surface wave are present. However, since the plasma beta is small, the slow mode disappears. For increasing values of G, the normalized phase speed of these waves increases as a function of ka. This situation is true when there is no flow. However, when flow is introduced, the phase speed gets significantly reduced. This shows that flows tend to dampen the phase speed of the fast Alfven gravity surface waves. It is interesting to note that the variation in the phase speed is significant only for $ka \approx 1$, while for $ka > 1$, the speed asymptotically approaches the phase speed of the body wave.

6.7 Negative Energy Waves

Plasma structures in nature show the presence of steady flows of the matter, which in general is directed along the direction of the magnetic field. However, in the Sun, there are down-flows along the photospheric magnetic flux tubes due to the presence of granules and super-granules, and spicules in the chromosphere. The presence of homogeneous steady flows generally leads to the Doppler shift in the wave frequencies. In the presence of a shift in the steady flow, there arises what are called the Kelvin−Helmholtz instabilities, of which we have more to say in a later chapter. It has been found that the presence of the transversal shift in the steady flows can lead to the appearance of backward waves. Slow steady flow velocities can change the dispersive characteristics of the MHD modes in magnetic structures.

An interesting observation about the general theory of waves is that backward waves (retarded) may have negative energy. The implication is that these waves tend to grow when the total energy of the system is decreased. The amplification of these waves can occur when there is dissipation in the system. This does not lead to any contradiction of the law of conservation of energy. The negative energy waves in a system may act as an efficient mechanism for the wave-flow interaction. The concept of negative energy waves has been studied in magnetic flux tubes with steady flow, where the kink modes with long wavelength limit were considered. For incompressible flows, an instability due to surface magneto sonic wave resonant absorption in the presence of steady flows has been interpreted as due to the presence of negative energy phenomena. This was discussed in chapter 5.

The presence of an inhomogeneity in the steady flow across a slab may result in the appearance of backward modes. There are trapped magneto sonic modes in the slab, which propagate in both directions (positive and negative) of the axis of the slab. The main criterion for the appearance of a backward wave can be defined as follows: waveguide system will have backward modes

© 2015 by Taylor & Francis Group, LLC

if the difference in the speeds of the external flows exceeds the phase speed of the mode considered, in the absence of steady flows.

The criterion for the presence of the negative energy waves in a waveguide system is that

$$C = \omega \frac{\partial D}{\partial \omega} < 0 \tag{6.69}$$

where $D(\omega, k_0)$ is a function evaluated at a particular value k_0 of the wavenumber k. The function D is related to the dispersion relation of the mode, by setting $D = 0$. An alternate formula for the presence of negative energy waves is $C = k\partial D/\partial k < 0$. More details about negative wave energy is found in Joarder et al[53].

6.8 Waves in Cylindrical Geometries

The appearance of many magnetic structures which are cylindrical flux tube shape in low β plasmas in magnetosphere and corona in the Sun encourages one to study the waves in cylindrical geometries (see Figure 6.11). The equations are similar to those of the slab geometries, except that the solutions of the wave equations are not in terms of hypergeometric functions, but in terms of the cylinder functions (Bessel functions) . Moreover, the cylindrical geometry brings in the curvature effects on the characteristics of these MHD waves.

Magnetic flux tubes act as wave guides allowing waves to propagate without spatial attenuation. Thus, they act as a good communication channel between one region of plasma and another. In much the same way, the loops of magnetic fields in the coronal atmosphere may act as communication channels. The flux tubes support a variety of MHD waves such as the Alfven waves, the magneto sonic waves. Sunspots support a variety of wave phenomenon that may be interpreted in terms of MHD waves The presence of MHD waves in the solar corona was just a theoretical suggestion until recently.

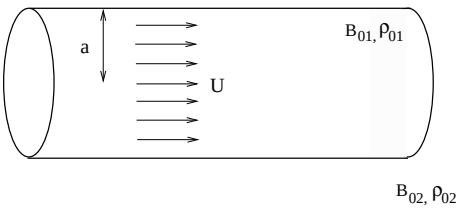


FIGURE 6.11: Cylindrical geometry with uniform flow inside the tube, from Satya Narayanan, Ramesh, Kathiravan, and Ebenezer [14].

© 2015 by Taylor & Francis Group, LLC

The evidence for the occurrence of magnetic flux tubes in astrophysical phenomena is increasing all the time. Flux tubes (or flux ropes) are believed to occur in the jets in some of the extragalactic sources, in the magnetospheres of some planets, and in the Sun. Indeed in the case of the Sun almost all of the magnetic flux emerging through the photospheric surface is found to occur in concentrated forms, ranging in scale from the visible sunspot to the very small ≈ 1" intense flux tubes. Thus, the source of coronal or solar wind magnetism is to be found in concentrated roots of the magnetic field emerging from the deep interior. First, we note that a tube is a wave guide. It permits waves to propagate without spatial attenuation. Thus, a tube is likely to provide a good communication channel between one region of a plasma and another, perhaps providing a connection between an energy source and an energy sink. For example, photospheric flux tubes connect the convection zone−an ample source of energy, especially in granules with the chromosphere and the corona. Secondly, we observe that a magnetic flux tube is an elastic object (and so an elastic, not rigid, wave guide) and as such is likely to respond to sudden changes by guiding waves. Sunspots are known to support a wide variety of wave phenomena. From an observational point of view the clearest evidence for flux tubes as magnetically distinct structures exists in the solar photosphere.

6.8.1 Different Types of Modes in Cylindrical Geometry

We consider the modes of oscillations of an isolated magnetic flux tube of strength B_0 embedded in a field free of pressure p_e and density ρ_e. If such a tube is subjected to a sudden twisting motion, it will respond by producing torsional oscillations of the tube. Such torsional oscillations may propagate as Alfven waves, for a tube of density ρ_0 and magnetic permeability μ. Compression of the gas is characterized by the sound speed:

$$c_{s0} = \frac{\gamma p_0}{\rho_0}^{1/2}$$

where γ is the ratio of the specific heats and p_0 is the gas pressure inside the tube. While the magnetic compressions are represented by Alfven speed V_A, the two speeds c_s and V_A combine to form

$$\left\{\frac{1}{c_s^2} + \frac{1}{V_A^2}\right\}^{1/2} = \frac{1}{c_T}$$

to produce the basic speed c_T. The tube speed c_T is both sub-Alfvenic and subsonic and, therefore, relates to the slow magneto acoustic wave. A symmetric squeezing of the tube may produce the sausage mode. Here, both gas and the magnetic field within the tube are expanding and contracting in the motion.

Asymmetric disturbances of the tube produce kink modes. This is similar to the waves on an elastic string, producing a propagation speed of $(T/\rho)^{1/2}$.

© 2015 by Taylor & Francis Group, LLC

The tension T in the string is clearly due to the magnetic tension, so $T = B_0^2/\mu_0$ and the density ρ is taken to be the sum of the gas density within the tube (ρ_0) and in the environment (ρ_e), i.e., $\rho = \rho_0 + \rho_e$.

The kink speed becomes

$$\begin{aligned} c_k &= \frac{B_0}{[\mu_0(\rho_0+\rho_e)]^{1/2}} = \frac{B_0}{[\mu_0\rho_0+\mu_0\rho_e]^{1/2}} \\ &= \frac{1}{(\mu_0\rho_0)^{1/2}/B_0 + (\mu_0\rho_e)^{1/2}/B_0} \end{aligned} \tag{6.70}$$

We have $V_A = B_0/(\mu_0\rho_0)^{1/2}$. Thus

$$c_k = V_A \left\{ \frac{\rho_0}{\rho_0+\rho_e} \right\}^{1/2}$$

The kink speed c_k is sub-Alfvenic but not necessarily subsonic. Additionally the radial behavior of the amplitude within $r \leq a$ may be either oscillating or decaying. Oscillating modes are classified as body waves while decaying or evanescent modes are surface waves.

We consider a uniform cylinder of magnetic field, $B_0\hat{z}$, confined to a radius of a surrounded by a magnetic field, $B_e\hat{z}$. The gas pressure and density within the cylinder are p_0 and ρ_0 and outside p_e and ρ_e (Figure 6.11).

Pressure balance results in

$$p_0 + \frac{B_0^2}{2\mu} = p_e + \frac{B_e^2}{2\mu}$$

where μ is the magnetic permeability.

The sound and Alfven speeds inside and outside the cylinder are given by

$$c_{s0} = (\gamma p_0/\rho_0)^{1/2} \quad \Rightarrow \quad p_0 = \rho_0 c_0^2/\gamma$$

$$V_{A0} = \frac{B_0}{(\mu\rho_0)^{1/2}} \quad \Rightarrow \quad B_0 = V_{A0}(\mu\rho_0)^{1/2}$$

$$c_e = (\gamma p_e/\rho_e)^{1/2} \quad \Rightarrow \quad p_e = \rho_e c_e^2/\gamma$$

$$V_{Ae} = \frac{B_e}{(\mu\rho_e)^{1/2}} \quad \Rightarrow \quad B_e = V_{Ae}(\mu\rho_e)^{1/2}$$

where γ is the ratio of specific heats.

Substituting the above equations we get the relation connecting the densities ρ_0 and ρ_e, i.e.,

$$\frac{\rho_0}{\rho_e} = \frac{2c_0^2 + \gamma V_{A0}^2}{2c_e^2 + \gamma V_{Ae}^2}$$

© 2015 by Taylor & Francis Group, LLC

Linear perturbations about this equilibrium lead to two equations:

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial^2}{\partial t^2} - (c_0^2 + V_A^2)\nabla^2\right)\triangle + c_{s0}^2 V_A^2 \frac{\partial^2}{\partial z^2}\nabla^2\triangle = 0 \quad (6.71)$$

$$\left(\frac{\partial^2}{\partial t^2} - V_A^2\frac{\partial^2}{\partial z^2}\right)\Gamma = 0 \quad (6.72)$$

where ∇^2 is the Laplacian operator in cylindrical coordinates, r, θ, z (equation B.15; see Appendix B1.4) and

$$\Gamma = \hat{z}\cdot\nabla\times V = \frac{1}{r}\frac{\partial}{\partial r}(rV_\theta) - \frac{1}{r}\frac{\partial v_r}{\partial\theta}$$

where $V = (v_r, v_\theta, v_z)$.

If $\triangle = R(r)\exp[i(\omega t + n\theta + kz)]$, then basic equations implie that R(r) satisfies the Bessel's equation given by

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} - \left(m_0^2 + \frac{n^2}{r^2}\right)R = 0 \quad (6.73)$$

where

$$m_0^2 = \frac{(k^2c_0^2 - \omega^2)(k^2V_A^2 - \omega^2)}{(c_0^2 + V_A^2)(k^2c_T^2 - \omega^2)} \quad (6.74)$$

For a solution bounded on the axis (r=0) of the cylinder:

$$R(r) = \begin{cases} A_0 I_n(m_0 r) & m_0^2 > 0, \\ A_0 J_n(m_0 r) & m_0^2 < 0 \end{cases}$$

where A_0 is a constant and I_n and J_n are Bessel functions of order n.

In the external region, with no propagation of energy away from R, the solution is given by

$$R(r) = A_1 K_n(m_e r), r > a$$

where

$$m_e = \frac{(k^2c_e^2 - \omega^2)(k^2V_{Ae}^2 - \omega^2)}{(c_e^2 + V_{Ae}^2)(k^2c_{Te}^2 - \omega^2)}$$

where

$$c_{Te}^2 = \frac{c_e^2 V_{Ae}^2}{c_e^2 + V_{Ae}^2}$$

Here m is taken to be positive.

Continuity of radial velocity component v_r and the total pressure across the cylinder (gas + magnetic) boundary (r = a) yields the required dispersion relations:

$$\rho_0(k^2V_{A0}^2 - \omega^2)m_e\frac{K_n'(m_e a)}{K_n(m_e a)} = \rho_e(k^2V_{Ae}^2 - \omega^2)m_0\frac{I_n'(m_0 a)}{I_n(m_0 a)} \quad (6.75)$$

© 2015 by Taylor & Francis Group, LLC

for surface waves $m_0^2 > 0$ and

$$\rho_0(k^2V_{A0}^2 - \omega^2)m_e\frac{K_n'(m_e a)}{K_n(m_e a)} = \rho_e(k^2V_{Ae}^2 - \omega^2)n_0\frac{J_n'(n_0 a)}{J_n(n_0 a)} \tag{6.76}$$

for body waves $m_0^2 = -n_0^2 < 0$.

Here $n = 0$ represents the cylindrically symmetric, while $n = 1$ represents the asymmetric kink mode. The effect of uniform flows on the characteristics of waves in flux tubes has been studied. Numerical solutions of the dispersion relation for the cylindrical geometry (equations 6.75 and 6.76) is presented in Figure 6.12. For more details on the solution of the dispersion relation, refer to Erdelyi[15]).

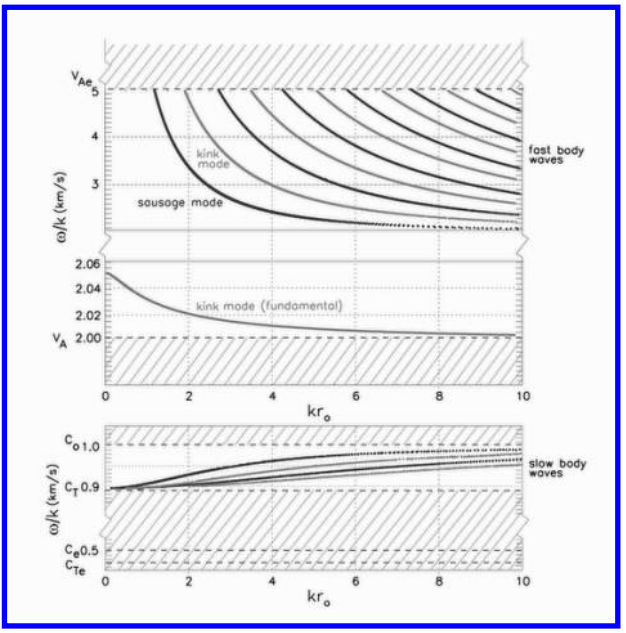


FIGURE 6.12: Solution of the dispersion relation for the cylindrical geometry, from Erdelyi [15].

6.9 Slender Flux Tube Equations

The slender flux tubes are typically tubes wherein the vertical motions slowly diverge with height z and have radii much smaller than the pressure scale height. The governing equations are the equation of continuity, vertical momentum, transverse momentum, and isentropic energy, given by

© 2015 by Taylor & Francis Group, LLC

$$\frac{\partial}{\partial t}\rho A + \frac{\partial}{\partial z}\rho \mathbf{v} A = 0 \tag{6.77}$$

$$\rho\Big[\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\frac{\partial \mathbf{v}}{\partial z}\Big] = -\frac{\partial p}{\partial z} - \rho g \tag{6.78}$$

$$p + \frac{B^2}{2\mu} = p_e \tag{6.79}$$

$$\frac{\partial p}{\partial t} + \mathbf{v}\frac{\partial p}{\partial z} = \frac{\gamma p}{\rho}\Big[\frac{\partial \rho}{\partial t} + \mathbf{v}\frac{\partial \rho}{\partial z}\Big] \tag{6.80}$$

$$BA = \text{constant} \tag{6.81}$$

The above equations govern the nonlinear behavior of longitudinal, isentropic motions $v(z,t)$ of a gas of density $\rho(z,t)$ and pressure $p(z,t)$ confined within an elastic tube of cross-sectional area $A(z,t)$. The derivation of the equations, as an expansion about the axis of the tube has been reported extensively in the literature. The case of the incompressible fluid has been considered. Also considered were isothermal and nonisothermal effects on the thin flux tube equations. By setting $v = 0$ and $\partial/\partial t = 0$, the equilibrium will be recovered. Assuming that the external medium is in hydrostatic equilibrium, with the same temperature and scale-height inside and outside the tube, the thin flux tube equations reduce to

$$p_0(z) = p_0(0)e^{-n(z)} \quad \rho_0(z) = \rho_0 \frac{\Lambda(0)}{\Lambda(z)} e^{-n(z)}$$

$$A_0(z) = A_0(0)e^{(1/z)n(z)}$$

where $n(z) = \int_0^z dz/\Lambda_0(z)$ and $\Lambda_0(z)$ is the pressure scale height. Linear perturbations of the equilibrium may be combined to yield (the details are skipped for brevity):

$$\frac{\partial^2 Q}{\partial t^2} - c^2(z)\frac{\partial^2 Q}{\partial z^2} + \omega_v^2(z)Q = 0 \tag{6.82}$$

where the speed $c(z)$ is defined by

$$\frac{1}{c^2} = \frac{1}{c_s^2(z)} + \frac{\rho_0(z)}{\Lambda_0(z)}\left[\frac{\partial A}{\partial p}\right]_{p=0} \tag{6.83}$$

and the frequency $\omega_v(z)$ by

$$\omega_v^2 = \omega_g^2 + c^2\left[\frac{1}{2}\Big[\frac{\rho_0'}{\rho_0} + \frac{A_0'}{A_0} + \frac{c'^2}{c^2}\Big]' + \Big[\frac{g}{c_s^2} - \frac{A_0'}{A_0}\Big]'\right.$$

$$\left.\frac{1}{4}\Big[\frac{rho_0'}{\rho_0} + \frac{A_0'}{A_0} + \frac{c'^2}{c^2}\Big]^2 + \Big[\frac{g}{c_s^2} - \frac{A_0'}{A_0}\Big]\Big[\frac{\rho_0'}{\rho_0} + \frac{c'^2}{c^2} + \frac{g}{c_s^2}\Big]\right] \tag{6.84}$$

© 2015 by Taylor & Francis Group, LLC

$Q(z,t)$ is related to the velocity $v(z,t)$ through

$$Q(z,t) = \left[\frac{\rho_0(z)\Lambda_0(z)c^2(z)}{\rho_0(0)\Lambda_0(0)c^2(0)}\right]^{1/2} v(z,t) \tag{6.85}$$

The evolution equation for $Q(z,t)$ is the Klein−Gordon equation which has been discussed for the symmetric sausage mode.

The analysis for the kink mode is slightly different. The linear transverse motions $v_\perp$ are governed by the equation:

$$\frac{\partial^2 v_\perp}{\partial t^2} = g\Big[\frac{\rho_0 - \rho_e}{\rho_0 + \rho_e}\Big]\frac{\partial v_\perp}{\partial z} + \Big[\frac{\rho_0}{\rho_0 + \rho_e}\Big]V_A^2\frac{\partial^2 v_\perp}{\partial z^2} \tag{6.86}$$

The first term on the right-hand side represents the buoyancy effects on the isolated flux tube, while the second term deals with the restoring force of the magnetic tension.

To get the Klein-Gordon equation for the kink mode, we write

$$v_\perp(z,t) = e^{z/4\Lambda_0}Q(z,t) \tag{6.87}$$

6.10 Waves in Untwisted and Twisted Tubes

6.10.1 Oscillations in Annular Magnetic Cylinders

Here, we shall consider a flux tube consisting of a central core surrounded by a shell or annulus layer, embedded in an uniform magnetic field. To begin with, we shall deal with an incompressible fluid wherein the phase speed of the slow and Alfven modes become indistinguishable. The fast modes are neglected. The longitudinal magnetic field in each of the regions will be as follows:

$$\mathbf{B} = \begin{cases} B_i = (0,0,B_i) & r < a, \\ B_0 = (0,0,B_0) & a \le r \le R, \\ B_e = (0,0,B_e) & r > R \end{cases}$$

where B_i, B_0, B_e are constant. The densities at the core, annulus, and external regions will be ρ_i, ρ_0 , and ρ_e, respectively. A similar expression for the pressure distribution will be assumed. The pressure balance conditions at the boundaries $r = a$ and $r = R$ are given by

$$p_i + \frac{B_i}{2\mu} = p_0 + \frac{B_0}{2\mu}$$

$$p_0 + \frac{B_0}{2\mu} = p_e + \frac{B_e}{2\mu}$$

© 2015 by Taylor & Francis Group, LLC

The distribution of the magnetic field and the annulus is shown in Figure 6.13.

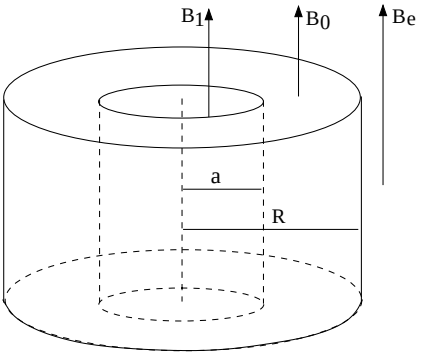


FIGURE 6.13: Distribution of the magnetic field in the annulus, from Carter and Erdelyi[16].

Assuming linear perturbations of the ideal MHD equations about the equilibrium and Fourier transforming the total Lagrangian pressure p_T and normal component of Lagrangian displacement ξ_r:

$$(p_T, \xi_r) \sim (\hat{p}_T(r), \hat{\xi}_r(r)) e^{i(m\theta + k_z z - \omega t}} \tag{6.88}$$

and omitting the hat of the Fourier decomposed perturbations for the sake of brevity, the total pressure satisfies the Bessel equation as given below:

$$\frac{d^2 p_T}{dr^2} + \frac{1}{r}\frac{dp_T}{dr} - \left(k_z^2 + \frac{m^2}{r^2}\right) p_T = 0 \tag{6.89}$$

where m is the azimuthal wavenumber. The dispersion relation after substituting the boundary conditions at boundaries $r = a$ and $r = R$, and some algebra reduces to

$$\frac{Q_0^i K_m'(k_z a) - (I_m'(k_z a) K_m(k_z a)/I_m(k_z a))}{I_m'(k_z a)(Q_0^i - 1)} = \frac{K_m'(k_z R)(Q_0^e - 1)}{Q_0^e I_m'(k_z R) - (K_m'(k_z R) I_m(k_z R)/K_m(k_z R))} \tag{6.90}$$

wherein

$$Q_0^i = \frac{\rho_i}{\rho_0}\frac{(\omega^2 - \omega_{Ai}^2)}{(\omega^2 - \omega_{A0}^2)} \tag{6.91}$$

$$Q_0^e = \frac{\rho_e}{\rho_0}\frac{(\omega^2 - \omega_{Ae}^2)}{(\omega^2 - \omega_{A0}^2)} \tag{6.92}$$

There are two modes (surface) to the above dispersion relation for each of

© 2015 by Taylor & Francis Group, LLC

the sausage and kink modes, respectively, for the annulus-core model. These modes propagate along the two natural surfaces of the system, i.e. at $r = a$ and $r = R$. The phase speeds get modified due to the annulus in comparison with a single straight tube, where the modes depend not only on the Alfven speed, but also on the ratio (a/R) of the core and annulus radii.

6.10.2 Magnetically Twisted Cylindrical Tube

Consider a flux tube embedded in a straight magnetic field (see Figure 6.14) given by

$$\mathbf{B} = \begin{cases} (0, Ar, B_0) & r < a \\ (0, 0, B_e) & r > a \end{cases}$$

The magnetic field and pressure, for the cylindrical equilibrium, satisfy the following equation:

$$\frac{d}{dr}\left[p_0 + \frac{B_{0\phi}^2 + B_{0z}^2}{2\mu}\right] + \frac{B_{0\phi}^2}{\mu r} = 0 \tag{6.93}$$

The second term in the above equation represents the magnetic pressure and the third term is due to the magnetic tension, derived due to the azimuthal component of the equilibrium magnetic field. Again, here we discuss only the incompressible case. We seek a bounded solution at $r = 0$ and $r \to \infty$. Applying the usual boundary conditions results in the following dispersion relation:

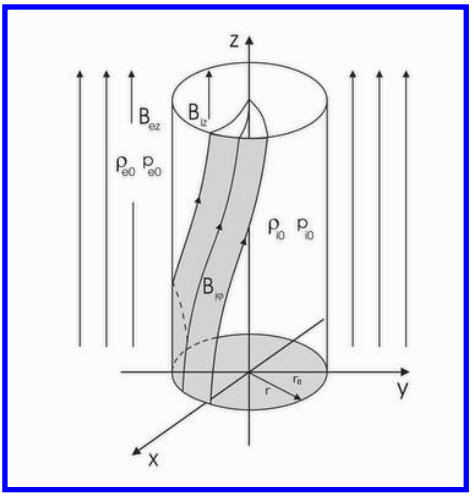


FIGURE 6.14: Distribution of the magnetic field in the twisted tube, from Erdelyi and Fedun[17].

© 2015 by Taylor & Francis Group, LLC

$$\frac{(\omega^2-\omega_{A0}^2)\dfrac{m_0 a I_m'(m_0 a)}{I_m(m_0 a)}-2m\omega_{A0}\dfrac{A}{\sqrt{\mu\rho_0}}}{(\omega^2-\omega_{A0}^2)^2-4\omega_{A0}^2\frac{A^2}{\mu\rho_0}} = \frac{|k_z|aK_m'(|k_z|a)}{K_m(|k_z|a)} \Bigg/ \left(\frac{\rho_e}{\rho_0}(\omega^2-\omega_{Ae}^2)+\frac{A^2}{\mu\rho_0}\frac{|k_z|aK_m'(|k_z|a)}{K_m(|k_z|a)}\right) \tag{6.94}$$

The dash in the above equation denotes derivative with respect to the argument of the Bessel function, and the frequencies are defined as

$$\omega_{A0}=\frac{1}{\sqrt{\mu\rho_0}}(mA+k_zB_0), \qquad \omega_{Ae}=\frac{k_zB_e}{\sqrt{\mu\rho_e}} \tag{6.95}$$

$$m_0^2=k_z^2\left[1-\frac{4A^2\omega_{A0}^2}{\mu\rho_0(\omega^2-\omega_{A0}^2)^2}\right] \tag{6.96}$$

The above equation 6.94 is the dispersion relation for MHD waves in an incompressible tube with uniform magnetic twist, embedded in a straight magnetic environment. No body waves exist for an incompressible fluid without magnetic twist. With a magnetic twist, a finite band of body waves arises. An interesting observation is that of the existence of a dual nature of the mode wherein a body wave exists for long wavelengths; however, surface wave characteristics are displayed for shorter wavelengths.

6.11 Applications to Coronal Waves

6.11.1 Kink Oscillations

The coronal imaging data movies clearly show the presence of fast-decaying quasi-periodic displacement of loops, often responding to an energy release nearby, in a form of a flare or eruption (flare generated oscillations). Analysis of 26 oscillating loops with lengths of 74−582 Mm, observed in EUV with TRACE (Aschwanden[54]) yielded periods 2.3−10.8 min, which is different for different loops. It is known that coronal loops are anchored in the dense plasma of the photosphere, so it is reasonable to assume that any motions in the corona are effectively zero at the base of a loop. The observed properties of these oscillations can be interpreted in terms of a kink fast magneto sonic mode (Edwin and Roberts[12]). The first observation of kink oscillations was after the flare on 14 July 1998 at 12.55 UT. The oscillation was identified as a global mode, with the maximum displacement situated near the loop

© 2015 by Taylor & Francis Group, LLC

apex and the nodes near the foot points. Using the above theory Nakariakov and Ofman[18] estimated the magnetic field in an oscillating loop as 13 ± 9 G. A typical sketch of the loop oscillation is presented in Figure 6.15 The effect of uniform flows on kink oscillations was studied by Satya Narayanan et al.[13] see also the review article by Satya Narayanan and Ramesh[55]. We briefly describe their results. Assume the plasma $\beta \ll 1$. The pressure balance condition is given by

$$p_0 + \frac{B_0^2}{2\mu} = p_e + \frac{B_e^2}{2\mu} \tag{6.97}$$

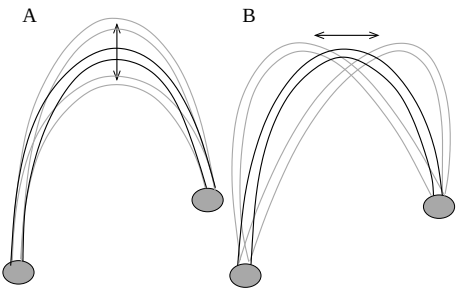


FIGURE 6.15: A loop oscillation, from Nakariakov and Offman[18].

For $\alpha = \rho_e/\rho_0$, $\epsilon = U_0/c_{A1}$, $x = \omega/kc_{A1}$, and low-β plasma, it can be shown that

$$m_e = k[1 - \alpha x^2]^{1/2} = m_e^* \tag{6.98}$$

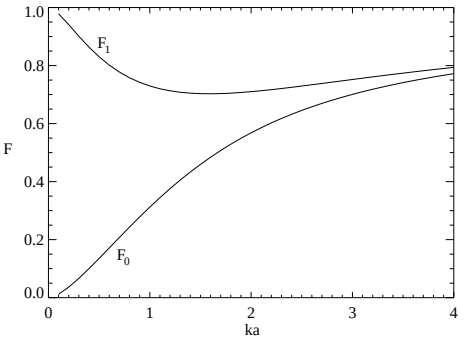


FIGURE 6.16: Behavior of $F_m(ka)$ for different values of ka, from Satya Narayanan, Ramesh, Kathiravan, and Ebenezer[14].

The dispersion relation for low-β plasma with flow can be written as

© 2015 by Taylor & Francis Group, LLC

$$[(x-\epsilon)^2-1](1-\alpha x^2)^{1/2}+\alpha(x^2-1)[1-(x-\epsilon)^2]^{1/2}F(m_0^*,m_e^*,a)=0 \quad (6.99)$$

$$F(m_0^*,m_e^*,a)=\frac{K_m(m_e^*a)I'_m(m_0^*a)}{K'_m(m_e^*a)I_m(m_0^*a)} \quad (6.100)$$

The behavior of $F_m(ka)$ for different values of ka is shown in Figure 6.16. The above relation (6.99) is highly transcendental and will have to be solved numerically. However, for $ka \ll 1$, one can show that $F(m_0^*, m_e^*, a) \approx 1$, so that the dispersion relation would reduce to

$$[(x-\epsilon)^2-1](1-\alpha x^2)^{1/2}+\alpha(x^2-1)[1-(x-\epsilon)^2]^{1/2}=0 \quad (6.101)$$

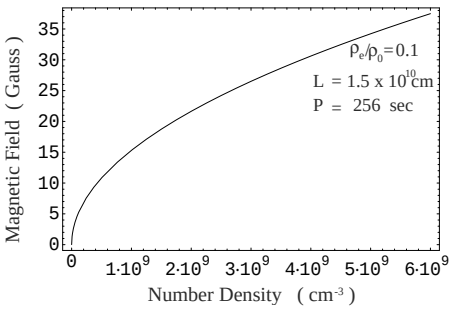


FIGURE 6.17: Variation of the magnetic field for different coronal parameters, from Satya Narayanan, Ramesh, Kathiravan, and Ebenezer[14].

The variation of the magnetic field for different coronal parameters is given in Figure 6.17.

For long wavelengths, the phase speed of the kink mode is about equal to the so-called kink speed c_K which, in the low-β plasma, is

$$c_K \approx \left[\frac{2}{1+n_e/n_0}\right]^{1/2} c_{A1} \quad (6.102)$$

where n_0 and n_e are the plasma concentrations inside and outside the loop, respectively, and c_{A1} is the Alfvén speed inside the loop.

It was shown by Nakariakov and Ofman [18] that the formula for the kink speed could be utilized to determine the magnetic field as follows:

$$B_0=(4\pi\rho_0)^{1/2}c_{A1}=\frac{\sqrt{2}\pi^{3/2}L}{P}\sqrt{\rho_0(1+\rho_e/\rho_0)} \quad (6.103)$$

© 2015 by Taylor & Francis Group, LLC

6.11.2 Sausage Oscillations

Modulated coronal radio emission which have periodicity in the range 0.5−5 s have been interpreted in terms of a fast magneto acoustic mode, the sausage mode, associated with the perturbations of the loop cross-section and plasma concentration by Nakariakov and Ofman[18]. Quasi periodic pulsations of shorter periods (0.5−10 s) may be associated with sausage modes of higher spatial harmonics (Roberts et al[56]; Nakariakov and Ofman[18].). There have been quasi−periodic pulsations in the periods 14−17 s, which oscillate in phase at a loop apex and its foot points which have been observed at radio wavelengths. These modes have a maximum magnetic field perturbation at loop apex and nodes and at the foot points. Dependence of the cut-off wavenumber of the sausage mode is depicted in Figure 6.18.

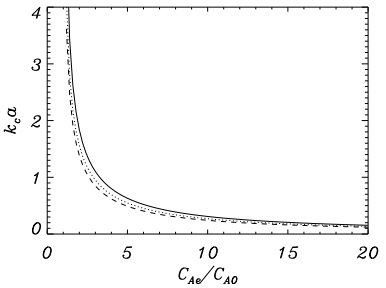


FIGURE 6.18: Dependence of the cut-off wavenumber of the sausage mode, from Nakariakov and Ofman [18].

The dispersion relation for magneto acoustic waves in cylindrical magnetic flux tubes has many types of long wavelength solutions in the fast mode branch ($n = 0, 1, 2, \ldots$) with the lowest case called the sausage ($n = 0$) and kink mode ($n = 1$). Kink mode solutions extend all the way to the long wavelength limit ($ka \to 0$) while the sausage mode has a cut−off at a phase speed of,

$$v_{ph} = v_{A2}$$

which has no solutions for wavenumbers $ka < k_c a$.

The cut-off wavenumber k_c is given by

$$k_c = \left[\frac{(c_s^2 + v_{A1}^2)(v_{A2}^2 - c_T^2)}{(v_{A2}^2 - v_{A1}^2)(v_{A2}^2 - c_s^2)}\right]^{1/2} \left(\frac{j_0}{a}\right)$$

Under coronal conditions the sound speed $c_s \approx 150 - 260$ km/s and Alfvén speed is $v_A \approx$ a few hundred km/s. Therefore,

$$c_s << v_A$$

© 2015 by Taylor & Francis Group, LLC

Here tube speed is similar to sound speed:

$$c_T \approx c_s$$

$$k_c \approx \left(\frac{j_0}{a}\right) \frac{1}{[(v_{A2}/v_{A1})^2 - 1]^{1/2}}$$

For a typical density ratio in the solar corona (0.1−0.5), the cut-off wavenumber $k_c a$ falls in the range $0.8 \leq k_c a \leq 2.4$. Therefore, the long wavelength sausage mode oscillation is completely suppressed for the slender loops. The occurrence of global sausage modes, therefore, requires special conditions: (1) very high density contrast ρ_0/ρ_e and (2) relative thick loops to satisfy $k > k_c$. The high density ratio $\rho_0/\rho_e >> 1$ or $v_{A2}/v_{A1} \gg 1$ yields the following simple expression for the cut-off wavenumber k_c:

$$k_c a \approx j_0(v_{A1}/v_{A2}) = j_0(\rho_e/\rho_0)^{1/2}$$

The cut-off wavenumber condition $k > k_c$ implies a constraint between the loop geometry ratio (2a/L) and the density contrast ratio (ρ_e/ρ_0) which turns out to be

$$\frac{L}{2a} \approx 0.65\sqrt{\rho_0/\rho_e}$$

Also it can be shown that the period of the sausage mode satisfies the condition

$$P_{saus} < \frac{2\pi a}{j_0 v_{A1}} \approx \frac{2.62a}{v_{A1}}$$

6.12 Nonlinear Aspects

In what follows, we briefly describe the equations of motion for slow solitary waves in multilayered magnetic structures. The basic model consists of a central magnetic slab sandwiched between two identical magnetic slabs (the equilibrium quantities are different from those in the central slab). This leads to certain tangential discontinuities in the flow properties. The two slabs surrounding the central slab are made up of two semi-infinite regions. To begin with, we shall discuss the dispersion relation for the linear case. The nonlinear evolution equation under the assumption that the characteristic wavelength of the central slab is much larger than the slab thickness is derived. Solitary-type solutions are obtained in this case which is slightly different from the

© 2015 by Taylor & Francis Group, LLC

Benjamin-Ono equation. The basic equations of motion are given by

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0 \tag{6.104}$$

$$\frac{\partial v}{\partial t} + (v \cdot \nabla)V = -\frac{1}{\rho}\nabla p + \frac{1}{\mu\rho}(\nabla \times B) \times B \tag{6.105}$$

$$\frac{\partial B}{\partial t} = \nabla \times (v \times B) \tag{6.106}$$

$$\nabla \cdot B = 0 \tag{6.107}$$

$$\frac{\partial}{\partial t}\left(\frac{p}{\rho^\gamma}\right) + v \cdot \nabla \left(\frac{p}{\rho^\gamma}\right) = 0 \tag{6.108}$$

Assuming linear approximation wherein the amplitude and phase of the waves have the form $\exp[i(kz - \omega t)]$, with real wavenumber k and frequency ω, the linear equations of motion can be reduced to two equations for u, the velocity, and total pressure $P' = p' + B_0 b_z/\mu$, as follows:

$$\frac{du}{dx} = \frac{i\omega(\omega^2 - c_s^2 k^2)}{\rho_0(c_s^2 + v_A^2)(\omega^2 - c_T^2 k^2)} P' \tag{6.109}$$

$$\frac{dP'}{dx} = \frac{i\rho_0(\omega^2 - v_A^2 k^2)}{\omega} u \tag{6.110}$$

where the squares of the Alfven, sound, and cusp speeds have already been defined earlier. Introducing suitable boundary conditions at the interfaces (continuity of total pressure and velocity across the discontinuities), the dispersion relation can be written as

$$\frac{\rho_i m_0(v_{Ai}^2 k^2 - \omega^2)}{\rho_0 m_i(v_A^2 k^2 - \omega^2)} \tan h(m_0 a)$$

$$= -\frac{\rho_i m_e(v_{Ai}^2 k^2 - \omega^2) + \rho_e m_i(v_{Ae}^2 k^2 - \omega^2) \tan h(m_i L)}{\rho_e m_i(v_{Ae}^2 k^2 - \omega^2) + \rho_i m_e(v_{Ai}^2 k^2 - \omega^2) \tan h(m_e L)} \tag{6.111}$$

The straightforward solution of the above equation can be written as

$$\omega = c_R k + k|k|\phi(k) \tag{6.112}$$

where

$$\phi(k) = \beta_\alpha \frac{1 - \lambda \exp(-2\kappa_\alpha |k| L)}{1 + \lambda \exp(-2\kappa_\alpha |k| L)}$$

$$\beta_\alpha = \frac{a\rho_\alpha c_T^3 (v_{A\alpha}^2 - c_T^2)}{2\rho_0 \kappa_\alpha v_A^4}$$

$$\kappa_\alpha^2 = \frac{(c_{s\alpha}^2 - c_T^2)(v_{A\alpha}^2 - c)T^2)}{(c_{s\alpha}^2 + v_{A\alpha}^2)(c_{T\alpha}^2 - c_T^2)}$$

$$\lambda = \frac{\rho_i \kappa_e (v_{Ai}^2 - c_T^2)\rho_e \kappa_i (v_{Ae}^2 - cT^2)}{\rho_i \kappa_e (v_{Ai}^2 - c_T^2) + \rho_e \kappa_e (v_{Ae}^2 - c_T^2)}$$

© 2015 by Taylor & Francis Group, LLC

In order to derive the nonlinear equation governing the propagation of slow surface sausage waves in the central magnetic slab, the method of reductive perturbation has been adopted. The nonlinear waves with amplitudes of the order of ϵ, propagate in the positive z direction with the phase velocity close to c_T. Utilizing expansions of the form

$$f = \epsilon f^{(1)} + \epsilon^2 f^{(2)} + \cdots \tag{6.113}$$

and substituting for the perturbations in the density, pressure, velocity, magnetic field, etc., in the higher order equations of the MHD equations, we obtain the equation

$$\frac{\partial \eta}{\partial t} + c_T \frac{\partial \eta}{\partial z} + q\eta \frac{\partial \eta}{\partial z} - \beta_i \mathcal{H}\left(\frac{\partial^2 \eta}{\partial z^2}\right) - \beta_i \mathcal{L}\left(\frac{\partial^2 \eta}{\partial z^2}\right) = 0 \tag{6.114}$$

where the Hilbert transform, $\mathcal{H}$, is given by

$$\mathcal{H}(f) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \frac{f(s)ds}{s-\theta}$$

$\mathcal{P}$ is the principal Cauchy part of an integral. The operator $\mathcal{L}$ is determined by

$$\mathcal{L}(f) = \int_{-\infty}^{\infty} G(z-s)f(s)ds$$

$$G(s) = \frac{\lambda}{\pi L}\int_0^{\infty} \frac{sin(rz/2L)}{\lambda + e^{r\kappa_i}}dr$$

Equation (6.114) describes nonlinear, weakly dispersive behavior of slow magneto sonic sausage waves. In the limit $L \to \infty$, $\mathcal{L}(f) \to 0$, the above equation reduces to the Benjamin−Ono equation. More details of the above equation, its derivation and the solitary wave solution can be found in Ruderman et al[57]. The discussion is rather technical and is skipped for brevity.

Theoretical models of the propagation of longitudinal waves in stratified coronal structures (see Nakariakov and Verwichte[58] for a very good review) describe the evolution of the wave shape and amplitude with the distance along the structures in terms of the extended Burger's equation given by

$$\frac{\partial A}{\partial s} - \alpha_1 A - \alpha_2 \frac{\partial^2 A}{\partial \xi^2} + \alpha_3 A \frac{\partial A}{\partial \xi} = 0 \tag{6.115}$$

where the coefficients α_1, α_2, and α_3 are, in general, functions of s. α_1 is a measure of the stratification, radiative losses, and heating, α_2 describes dissipation by thermal conductivity and viscosity, while α_3 is a measure of the nonlinearity. The running coordinate $\xi = s - c_s t$.

In this chapter, we have discussed the different types of MHD modes for nonuniform media. These waves are found in addition to the standard MHD modes the Alfven and the two magneto acoustic modes in an uniform media.

© 2015 by Taylor & Francis Group, LLC

Most of the waves that occur in nature are formed in nonuniform media. Thus, the discussion of the different modes in the present chapter will have interesting applications. We have discussed the importance of waves such as the kink and sausage modes in describing the plasma parameters in coronal loops. Finally, we have briefly talked about the nonlinear structure of the evolution equation for the slow sausage surface waves in multilayered magnetic structures. In the next chapter we shall deal with different aspects of shock waves in a compressible medium. In this case, the discontinuities are part and parcel of the medium, which propagate with and without dissipation. More discussion will follow in the next chapter.

6.13 Exercises

1. Given a tangential discontinuity in the magnetic fields and density, say $B_1 \neq B_2$ and $\rho_1 \neq \rho_2$, for a plasma whose β is small, discuss the Alfven velocities on either side of the medium.

2. The dispersion relation for interfacial (surface) waves is given by

$$\omega^2 = k_z^2[\frac{\rho_1 v_{A1}^2 + \rho_2 v_{A2}^2}{\rho_1 + \rho_2}]$$

   Derive the bound for the phase velocity in terms of the Alfven velocities.

3. In the presence of a uniform shear U, discuss the bound for the phase velocity of the surface wave. In the absence of the shear flow, show that this result is the same as in Exercise 2.

4. The dispersion relation for waves in a single tangential discontinuity for low β plasma is

$$\alpha x^4 + (1+\alpha)x^2 + cos^2\theta(1 + sin^2\theta) = 0$$

   $\alpha = B_{02}/B_{01}$. Calculate the phase speed $x = \omega\ kv_A$ for $\alpha = 0.8$ and $\theta = 0, 45°$.

5. The dielectric constants $\tau_{1,2}$ for the interfacial waves is given by

$$\tau_{1,2}^2 = \frac{\omega^4 - k^2 \sin^2(\theta + \gamma_{1,2})(V_{Ae,A}^2 + c_{se,s0}^2) + k^4 c_{se,s0}^2 V_{Ae,A}^2 \sin^4(\theta + \gamma_{1,2})}{k^2 c_{se,s0}^2 V_{Ae,A}^2 \sin^2(\theta + \gamma_{1,2}) - \omega^2(c_{se,s0}^2 + V_{Ae,A}^2)}$$

   Show that for the incompressible limit, $\tau_{1,2} \to k$,

© 2015 by Taylor & Francis Group, LLC

6. In the section on two-mode structure of Alfven surface waves, show analytically that the dispersion relation reduces to the nonviscous case when $\nu = 0$.

7. In a magnetically twisted (uniform) tube, the basic magnetic field has the form

$$B = \begin{cases} (0, Ar, B_0), & r < a \\ (0, 0, B_e), & r > a \end{cases}$$

   Considering the effect of flows in this geometry within the cylinder, what should be the form of the basic shear and explain why?

8. If

$$F(m_0^*, m_e^*, a) = \frac{K_m(m_e^* a) I'_m(m_0^* a)}{K'_m(m_e^* a) I_m(m_0^* a)},$$

   show that for $k << 1$, $F(m_0^*, m_e^*, a) \approx 1$.

9. Calculate the magnetic field from the relation derived for kink oscillations if $\rho_e/\rho_0 = 0.2$, $L = 2 \times 10^{11}$ cm, $P = 164$ sec and $\rho_0 = 4 \times 10^9/\text{cm}^3$.

10. Calculate the Alfven velocity v_{A1} if the radius of the flux tube is $500km$ and the period of the sausage oscillation is 4 sec .

© 2015 by Taylor & Francis Group, LLC

Chapter 7

Shock Waves

7.1 Introduction 311
7.2 Discontinuities in Surfaces 312
7.3 Normal Shock Waves 313
7.4 Oblique Shock Waves 315
7.5 Blast Waves – Similarity Solution of Taylor–Sedov 316
7.6 Weak Shock Waves 319
7.6.1 Example of Weak Shocks 322
7.7 Waves in a Polytropic Gas 323
7.8 An Application of Shock Waves in the Sun 326
7.8.1 An Example from Earth's Bow Shock 328
7.9 Shock Waves in Collisionless Plasmas 328
7.9.1 Dispersive Shock Waves 331
7.10 Shocks in MHD 335
7.10.1 Parallel Shocks 337
7.10.2 Perpendicular Shocks 337
7.11 Nonlinear Studies 341
7.11.1 Burger's Equation 343
7.11.2 Stationary Solutions and Shock Structure 347
7.11.3 Single-Hump Solution 348
7.11.4 Planar N-Wave 349
7.11.5 Backlund Transformation 351
7.12 Exercises 351

7.1 Introduction

In the previous chapters on hydrodynamics and magnetohydrodynamics (MHD), we covered wave propagation in the presence of tangential discontinuities in the magnetic fields, presence of flows, etc. In the studies mentioned above, the corresponding dispersion relation was derived with the assumption that the total pressure and the vertical component of the velocity field are continuous across the tangential discontinuity. In the case of the fluid or plasma being compressible, the velocity of sound plays an important role. There are situations in several branches of physics, such as MHD, space plasmas, and gas dynamics in which the pressure develops a discontinuity, and in the absence of dissipative effects such as viscosity and heat conductivity, propagation and convection of compressional disturbances, with speeds increasing with compression, lead to a continual steepening of waveforms that eventually can no longer be expressed by single-valued functions of position. In such cases, we introduce a discontinuity in the flow to get over this difficulty. A shock wave

© 2015 by Taylor & Francis Group, LLC

may be defined as a surface in a flow field across which the flow variables change discontinuously. It is important that shock waves exist in order that one may admit certain types of boundary conditions that could not be satisfied in a continuous flow pattern. In practical situations, a real fluid cannot sustain an actual discontinuity so the latter is only an idealization of the sharp gradients in the flow variables that occur in reality in a shock wave. A typical situation that one encounters in everyday life is the sonic boom that is heard at the wake of a supersonic aircraft with a Mach number larger than unity. The Mach number is a nondimensional number which relates to the ratio of the flow speed to the sound speed. For $M > 1$, the flow is supersonic, However, for $M < 1$ and $M \approx 1$, the flows are termed as subsonic and transonic, respectively. In this chapter, we shall study different types of shocks, such as normal and oblique shocks. We shall briefly discuss weak shocks, waves in a polytropic gas and relaxing medium. We will also briefly mention the Sedov blast wave solution. The Burger's equation and the KdV equation in shocks will also be dealt with, to include nonlinear effects. Application of shocks in MHD and space plasmas will also be included.

7.2 Discontinuities in Surfaces

In most of the cases in nature, the flow parameters such as pressure, velocity, and density vary continuously. However, there are situations in which discontinuities in the distribution of these quantities occur. A discontinuity in a fluid flow occurs over one or more surfaces; the concerned quantities change discontinuously as one crosses such a surface. These surfaces maybe termed "surfaces of discontinuity". For unsteady flows, the surfaces of discontinuity do not in general remain fixed; also the rate of change of motion of these surfaces may not have any direct relation with the velocity of the flow. The particles in their motion may cross a surface of discontinuity. However, one has to invoke certain boundary conditions. We shall briefly discuss such boundary conditions.

It is important to make sure that the flux is continuous; i.e., the mass of gas coming from one side must equal the mass leaving the other side. If we define the mass flux through the surface element as ρv_x, then we must have the relation $\rho_1 v_{x1} = \rho_2 v_{x2}$, where the indices 1 and 2 denote the two sides of the surface of discontinuity. We can denote such a relation as

$$[\rho v_x] = 0 \tag{7.1}$$

The next important condition is the continuity of the energy flux. This can be written as

$$[\rho v_x (1/2)v^2 + \mathcal{P}] = 0 \tag{7.2}$$

© 2015 by Taylor & Francis Group, LLC

where $\mathcal{P}$ is a measure of the pressure gradient.

Finally, one more relation with respect to the momentum flux should also be introduced to compensate the boundary conditions. This is given by

$$[p + \rho v_x^2] = 0 \tag{7.3}$$

and the y and z components relations are given by

$$[\rho v_x v_y] = 0 \qquad [\rho v_x v_z] = 0 \tag{7.4}$$

Equations (7.2−7.4) form a complete system of boundary conditions at a surface of discontinuity. We can think of two types of surfaces of discontinuity. In the first, no mass flux passes through the surface while in the second the mass flux is nonzero. From the first type, we can infer that the normal velocity component and the gas pressure are continuous at the surface of discontinuity, while the tangential velocities v_y and v_z and the density may be continuous. Such a discontinuity may be called a tangential discontinuity. Mathematically such a relation can be written as

$$v_{1x} = v_{2x} = 0, \qquad [p] = 0 \tag{7.5}$$

In the case when the mass flux is not zero, v_{1x} and v_{2x} are, therefore, not zero. This leads to the condition

$$[v_y] = 0, \qquad [v_z] = 0 \tag{7.6}$$

In summary, the boundary conditions can be written as (same as equations (7.1−7.3)

$$[\rho v_x] = 0 \tag{7.7}$$

$$[(1/2)v_x^2 + \mathcal{P}] = 0 \tag{7.8}$$

$$[p + \rho v_x^2] = 0 \tag{7.9}$$

A discontinuity satisfying the above system of conditions which propagates in a compressible medium is defined as a shock wave or simply called a shock.

7.3 Normal Shock Waves

It is well known that the changes in the flow properties that occur across a shock wave may be determined without actually knowing the details on the dissipation processes which happen within the shock wave. In general, the dissipation processes influence mostly the structure within the shock wave and not the flow changes as mentioned above. We shall discuss a shock wave with

© 2015 by Taylor & Francis Group, LLC

its plane normal to the flow. The equations of motion (one-dimensional) may be written as[164]:

$$\frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial x} + \rho\frac{\partial u}{\partial x} = 0 \tag{7.10}$$

$$\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right) + \frac{\partial p}{\partial x} = 0 \tag{7.11}$$

$$\frac{\partial S}{\partial t} + u\frac{\partial S}{\partial x} = 0 \tag{7.12}$$

where S is the entropy of the system. The above equations of motion can also be written in conservative form. We shall skip it for brevity. We shall introduce a transformation such that the frame of reference is moving the shock wave. Integrating the resulting equations across the shock wave, and using simple algebra, we obtain the following jump conditions:

$$-U[\rho] + [\rho u] = 0 \tag{7.13}$$

$$-U[\rho u] + [\rho u^2 + p] = 0 \tag{7.14}$$

$$-U[(1/2)\rho u^2 + \rho e] + [((1/2)\rho u^2 + \rho e)u + pu] = 0 \tag{7.15}$$

e is the internal energy of the system. Introducing the simple transformation

$$v = U - u \tag{7.16}$$

the above equations can be simplified to yield

$$[\rho v] = 0 \tag{7.17}$$

$$[p + \rho v^2 - \rho v U] = 0 \tag{7.18}$$

$$[\rho v(h + (1/2)v^2) - (p + \rho v^2)U + (1/2)\rho v U^2] = 0 \tag{7.19}$$

Here, h is the enthalpy of the system. With simple algebra, the above equations can be written as

$$\rho_1 V_1 = \rho_2 V_2 \tag{7.20}$$

$$p_1 + \rho_1 V_1^2 = p_2 + \rho V_2^2 \tag{7.21}$$

$$h_1 + (1/2)V_1^2 = h_2 + (1/2)V_2^2 \tag{7.22}$$

The subscripts 1 and 2 refer to the flow condition in front of and behind the shock. These relations in literature are called the famous Rankine−Hugoniot(RH) conditions. A typical Hugoniot curve is presented in Figure 7.1.

Using equations (7.20 and 7.21), we have

$$V_2^2 - V_1^2 = (p_1 - p_2)\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right) \tag{7.23}$$

© 2015 by Taylor & Francis Group, LLC

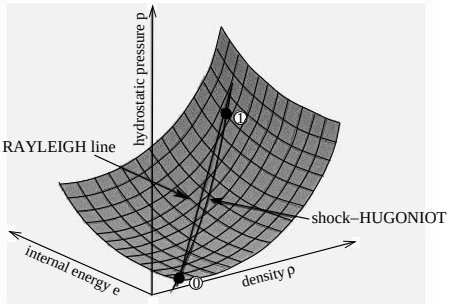


FIGURE 7.1: Hugoniot curve.

Using the above equation into equation (7.22), one gets

$$\frac{(p_1 - p_2)}{2}\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right) = \frac{\gamma}{\gamma - 1}\left(\frac{p_1}{\rho_1} - \frac{p_2}{\rho_2}\right) \tag{7.24}$$

Here, γ is the ratio of the specific heats. From the above relations, the ratios of the pressure and the density can be written as follows:

$$\frac{p_2}{p_1} = \frac{\left(\frac{\gamma+1}{\gamma-1}\right)\frac{\rho_2}{\rho_1} - 1}{\left(\frac{\gamma+1}{\gamma-1}\right) - \frac{\rho_2}{\rho_1}} \tag{7.25}$$

A similar expression for the ratio of the densities can be written as

$$\frac{p_2 - p_1}{\rho_2 - \rho_1} = \gamma\frac{p_2 + p_1}{\rho_2 + \rho_1} \tag{7.26}$$

In terms of the Mach numbers M_1 and M_2, the ratios of the pressure and the densities can be written as

$$\frac{p_2}{p_1} = \frac{1 + \gamma M_1^2}{1 + \gamma M_2^2} \tag{7.27}$$

and

$$\frac{\rho_2}{\rho_1} = \frac{\left(\frac{\gamma+1}{2}\right)M_1^2}{1 + \left(\frac{\gamma-1}{2}\right)M_1^2} \tag{7.28}$$

7.4 Oblique Shock Waves

In this section, we shall consider a shock wave whose plane is at an angle (oblique) to the flow direction. The governing equations which deal with the

© 2015 by Taylor & Francis Group, LLC

conservation of mass, momenta, tangential, and normal to the shock and the energy are written as

$$\rho_1 V_{n1} = \rho_2 V_{n2} \tag{7.29}$$
$$(\rho_1 V_{n1}) V_{t1} = (\rho_2 V_{n2}) V_{t2} \tag{7.30}$$
$$p_1 + \rho_1 V_{n1}^2 = p_2 + \rho_2 V_{n2}^2 \tag{7.31}$$
$$h_1 + V_1^2/2 = h_2 + V_2^2/2 \tag{7.32}$$

The subscripts n and t denote the values perpendicular and parallel to the shock wave, respectively. From relation (30), it follows that

$$V_{t1} = V_{t2} \tag{7.33}$$

If we replace M_1 by $M_1 \sin\theta$, and simplify, the ratio of the pressure and density can be written as

$$\frac{p_2}{p_1} = 1 + \frac{2\gamma}{\gamma+1}(M_1^2 \sin^2\theta - 1) \tag{7.34}$$
$$\frac{\rho_2}{\rho_1} = \frac{\tan\theta}{\tan(\theta - \delta)} \tag{7.35}$$

where δ is given by

$$\tan\delta = \frac{(M_1^2 \sin^2\theta - 1)\cot\theta}{1 + M_1^2\left(\frac{\gamma+1}{2} - \sin^2\theta\right)} \tag{7.36}$$

7.5 Blast Waves – Similarity Solution of Taylor–Sedov

A discussion on the normal and oblique shocks was considered above. However, in what follows, we shall deal with a spherical shock (the propagation is with respect to the radial distance) produced by a strong explosion, i.e., an instantaneous release of a large amount of energy E in a small volume. Assume that the wave distance is small enough from the source so that its amplitude is reasonably large. Assuming the shock wave to be strong, one can neglect the pressure p_1 of the undisturbed stagnant gas in front of it in comparison with the pressure p_2 immediately behind it. The density ratio approaches the value given by

$$\frac{\rho_2}{\rho_1} \approx \frac{\gamma+1}{\gamma-1} \tag{7.37}$$

We can say that the flow pattern is essentially determined by two parameters, E and ρ_1, the energy and the density. We seek some self-similar solution of

© 2015 by Taylor & Francis Group, LLC

the gas dynamic set of equations (refer equation 7.3). For this we have to introduce a nondimensional parameter given by

$$\xi = R\left(\frac{\rho_1}{Et^2}\right)^{1/5} \tag{7.38}$$

Neglecting p_1, the expression for the energy is given by

$$E = \int_0^\infty \left[\frac{p}{\gamma - 1} + (1/2)\rho V^2\right] 4\pi R^2 dR \tag{7.39}$$

We also assume that p, ρ, and V are all functions of ξ only. The shock wave position corresponds to a certain constant ξ_0 of ξ. If $R = R_0$ denotes the shock radius at a given time t, then we can show that

$$R_0 = \xi_0\left(\frac{Et^2}{p_1}\right)^{1/2} \tag{7.40}$$

The above relation is useful to determine the rate of propagation of the shock wave which is given by

$$u_1 = \frac{dR_0}{dt} = \frac{2R_0}{5t} \tag{7.41}$$

We introduce the following nondimensional variables:

$$V = \frac{4}{5(\gamma+1)}\frac{R}{t}V'(\xi) \tag{7.42}$$

$$\rho = \frac{\gamma+1}{\gamma-1}\rho_1\rho'(\xi) \tag{7.43}$$

$$p = \frac{8\rho_1}{25(\gamma+1)}\frac{R^2}{t^2}p'(\xi) \tag{7.44}$$

The boundary conditions on the shock wave are given by

$$\xi = \xi_0; \qquad V', p', \rho' = 1 \tag{7.45}$$

The basic equations get transformed into three sets of ordinary differential equations given by

$$\xi[(V' - (2/5))\frac{dV'}{d\xi'} + \frac{p'}{\rho'}] = -V'(V'-1) - 2\frac{p'}{\rho'} \tag{7.46}$$

$$\xi[\frac{dV'}{d\xi} + (V' - (2/5))\frac{1}{\rho'}\frac{d\rho'}{d\xi}] = -3V' \tag{7.47}$$

$$\xi(V' - (2/5))(\frac{1}{p'}\frac{dp'}{d\xi} - \frac{\gamma}{\rho'}\frac{d\rho'}{d\xi}) = -2(V'-1) \tag{7.48}$$

© 2015 by Taylor & Francis Group, LLC

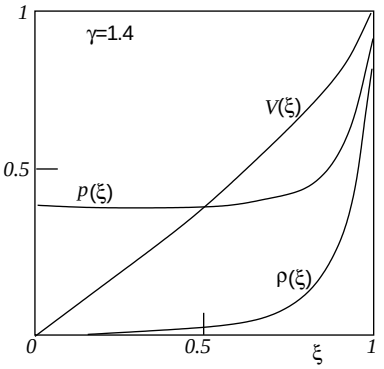


FIGURE 7.2: Velocity, pressure, and density (nondimensional) as a function of scaled radius behind the Sedov−Taylor blast wave.

The constancy of the total energy between any two similarity lines $R/t^{2/5} = constant$ leads to the relation

$$\left(\frac{p}{\gamma-1}+(1/2)\rho V^2\right)\left(V-\frac{2R}{5t}\right)+pV=0 \tag{7.49}$$

Taylor and Sedov gave an elegant solution of the above set of equations which is given by

$$\left(\frac{\xi_0}{\xi}\right)^5 = V'^2\left[\frac{5(\gamma+1)-2(3\gamma-1)V'}{7-\gamma}\right]^{\nu_1}\left[\frac{2\gamma V'-(\gamma+1)}{\gamma-1}\right]^{\nu_2} \tag{7.50}$$

$$\rho' = \left[\frac{2\gamma V'-(\gamma+1)}{\gamma-1}\right]^{\nu_3}\left[\frac{5(\gamma+1)-2(3\gamma-1)V'}{\gamma-1}\right]^{\nu_4}\left[\frac{\gamma+1-2V'}{\gamma-1}\right]^{\nu_5} \tag{7.51}$$

The ν are given by

$$\nu_1=\frac{13\gamma^2=7\gamma+12}{(3\gamma-1)(2\gamma+1)};\quad \nu_2=-\frac{5(\gamma-1)}{2\gamma+1}$$

$$\nu_3=\frac{3}{2\gamma+};\quad \nu_4=\frac{13\gamma^2-7\gamma+12}{(2-\gamma)(3\gamma-1)(2\gamma+1)};\quad \nu_5=\frac{1}{\gamma-2}$$

The constant ξ_0 can be calculated from the following relation:

$$\frac{32\pi\xi_0^2}{25(\gamma^2-1)}\int_0^1(\xi^4\rho' V'^2+\xi^9 p')d\xi=1 \tag{7.52}$$

The similarity solution obtained by Taylor and Sedov is presented in Figure 7.2. The asymptotic behavior is given by

$$\frac{R}{R_0}\Rightarrow 0;\quad \frac{V}{V_2}\approx\frac{R}{R_0};\quad \frac{\rho}{\rho_2}\approx\left(\frac{R}{R_0}\right)^{3/\gamma-1} \tag{7.53}$$

© 2015 by Taylor & Francis Group, LLC

7.6 Weak Shock Waves

A weak shock is defined as one in which the discontinuities in every quantity is very small. Let us briefly describe the shock theory of Whitham[174]. To begin with we shall discuss the continuous solution of a nonlinear equation, using the method of characteristics. Once we obtain such a solution, we shall move on to describe the weak solution of the shock. Other topics such as shock capturing, fitting, and related topics are more technical and involved and will not be dealt in this book.

Consider the first-order differential (partial) for the density ρ as follows:

$$\rho_t + c_0\rho_x = 0, \qquad c_0 = \text{constant} \tag{7.54}$$

The independent variables are x and t. The above equation is linear and so the general solution can be immediately written as

$$\rho = f(x - c_0 t)$$

where $f(x)$ may be an arbitrary function, while the solution of any particular situation may be achieved by simply matching the function f to an initial or boundary condition. The implication of the solution is that it describes a wave motion with the initial profile $f(x)$ being translated unchanged in shape a distance $c_0 t$ to the right at time t.

If the equation is linear, then, the analysis is straightforward. However, consider the nonlinear equation

$$\rho_t + c(\rho)\rho_x = 0 \tag{7.55}$$

where $c(\rho)$ is a given function of ρ. One of the approaches to the solution of the nonlinear equation is to consider the function $\rho(x,t)$ at each point of the (x,t) plane and that $\rho_t + c(\rho)\rho_x$ is the total derivative of ρ along a curve whose slope is given by

$$\frac{dx}{dt} = c(\rho) \tag{7.56}$$

at every point of it. In the (x,t) plane, for any curve, x and ρ may be considered to be functions of t, with the total derivative of ρ as

$$\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \frac{dx}{dt}\frac{\partial\rho}{\partial x}$$

Consider a curve $\mathcal{C}$ in the $(x.t)$ plane which satisfies equation (7.56). From the above relation, one can deduce that

$$\frac{d\rho}{dt} = 0. \qquad \frac{dx}{dt} = c(\rho) \tag{7.57}$$

© 2015 by Taylor & Francis Group, LLC

Consider the initial value problem:

$$\rho = f(x), \qquad t = 0, \qquad -\infty < x < \infty \tag{7.58}$$

If one of the curves $\mathcal{C}$ intersects $t = 0$ at $x = \xi$, then $\rho = f(\xi)$ on the curve. The equation for the curve (see Figure 7.3) is

$$x = \xi + F(\xi)t$$

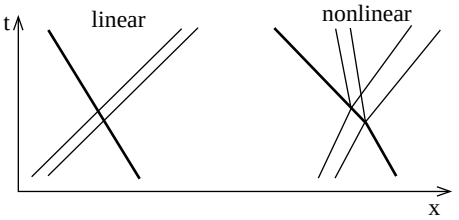


FIGURE 7.3: Left: for the linear hyperbolic system, the solution is a simple superposition of traveling waves. Right: for the nonlinear system, waves of different families have nontrivial interactions.

Allowing ξ to vary, one obtains the whole family of solutions:

$$\rho = f(\xi). \qquad c = F(\xi) = c(f(\xi)) \tag{7.59}$$

on

$$x = \xi + F(\xi)t \tag{7.60}$$

From the expressions (7.59) and (7.60), we have

$$\rho_t = f'(\xi)\xi_t, \qquad \rho_x = f'(\xi)\xi_x$$

and from the derivatives of (7.60) in t and x, we have

$$F(\xi) + [1 + F'(\xi)t]\xi_t = 0$$

$$[1 + F'(\xi)t]\xi_x = 1$$

Thus,

$$\rho_t = -\frac{F(\xi)f'(\xi)}{1 + F'(\xi)t}, \qquad \rho_x = \frac{f'(\xi)}{1 + F'(\xi)t} \tag{7.61}$$

From the above expression, we see that

$$\rho_t + c(\rho)\rho_x = 0$$

is automatically satisfied. The curves used in constructing the solution of the above equation are called the characteristic curves. One encounters situations

© 2015 by Taylor & Francis Group, LLC

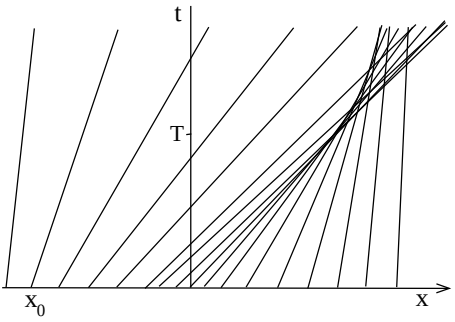


FIGURE 7.4: After a time, T, when the characteristics start to intersect, a shock is produced.

when there are discontinuities in the density, pressure, etc. Solutions which incorporate such discontinuities are called shock solutions (see Figure 7.4). We have mentioned their behaviors in the previous sections.
It is interesting to note that there are a number of situations wherein the shocks are weak in that $(\rho_2 - \rho_1)/\rho_1$ is rather small. However, they may not be extremely weak that they are treated as discontinuities. One resorts to some approximations on such cases. For example. the shock velocity given by

$$V = \frac{Q(\rho_2) - Q(\rho_1)}{\rho_2 - \rho_1}$$

may tend to the characteristic velocity given by

$$c(\rho) = \frac{dQ}{d\rho}$$

in the limit of the shock strength $(\rho_2 - \rho_1)/\rho_1 \to 0$. In order to incorporate weak shocks, one expands the shock velocity V in a Taylor series in $(\rho_2 - \rho_1)/\rho_1$ as

$$V = Q'(\rho_1) + (1/2)(\rho_2 - \rho_1)Q"(\rho_1) + O(\rho_2 - \rho_1)^2$$

In a similar way, the propagation velocity $c(\rho_2) = Q'(\rho_2)$ may be expanded as

$$c(\rho_2) = c(\rho_1) + (\rho_2 - \rho_1)Q"(\rho_1) + O(\rho_2 - \rho_1)^2$$

Thus,

$$V = (1/2)(c_1 + c_2) + O(\rho_2 - \rho_1)^2 \tag{7.62}$$

where $c_1 = c(\rho_1)$ and $c_2 = c(\rho_2)$. The shock velocity is the mean of the characteristic velocities on the two sides of it. In the (x, t) plane, the shock curve bisects the angle between the characteristics which meet on the shock. An interesting point is that this relation is exact when $Q(\rho)$ is a quadratic function. In passing, it would be worth making the statement that a continuous

© 2015 by Taylor & Francis Group, LLC

wave breaks and produces a shock if and only if the propagation velocity c decreases as x increases. In such a scenario,

$$c_2 > V > c_1$$

7.6.1 Example of Weak Shocks

In this example. let us consider a conservation law from fluid dynamics. This equation is governed by the inviscid Burger's equation and is applied to modeling gas dynamics. Burger's equation with dissipation will be dealt with in greater detail towards the end of the chapter. It is a typical equation for which the solution can develop discontinuities (shock waves).

Consider the following equation with boundary conditions:

$$u_t + uu_x = 0$$

$$u(x, 0) = f(x)$$

The above problem is a typical initial value problem. This has a simple implicit solution of the form

$$u(x, t) = f(x - ut)$$

with the constraint that $1 + tf'(x - \xi t) \neq 0$ for $t \geq 0$. Thus, if $f'(x - \xi t)$ is negative then it will vanish at $t = 1/f'$.

An alternate procedure is to get back to the integral form of the conservation law and look for a weak solution of this equation. The general form of a conservation law may stated as follows: the change in time of a quantity such as mass, charge, and momentum for which a density function $u(x, t)$ is due to the flux of that quantity on the boundary, given by a flux function F evaluated on α and x. Here α is a given constant. The rate of change of the density function satisfies the relation

$$\frac{d}{dt}\int_{\alpha}^{x} u(y, t)dy = F(u(x, t)) - F(u(\alpha, t))$$

If u is a regular function, then we can push the differential sign inside and use integration by parts to obtain the partial differential equation

$$u_t - F'(u)u_x = 0$$

Comparing the above relation with the original equation, we have $F(u) = -u^2/2$. We invoke the Rankine−Hugoniot jump conditions given by

$$\frac{ds}{dt} = \frac{F(u_+) - F(u_-)}{u_+ - u_-}$$

where $u_+(x_0)$ and $u_-(x_0)$ are the limits when $x \to x_0$ from either side of $x = s(t)$.

© 2015 by Taylor & Francis Group, LLC

It is interesting to note that a weak solution, defined by the above conditions, satisfies the integral form of the corresponding conservation law. We shall sketch a simple proof for the above statement. It is trivial to realize that in those regions which do not have discontinuities, the PDE (partial differential equation) and the integral equation are equivalent and there is nothing much to prove.

Consider a small circle Ω centered around a point $(s(t), t)$. Denote by Ω_+ the fraction on the right side, and Ω_- on the left. For all $\phi \in D(\Omega)$,

$$\int_{\Omega_+} u\phi_t - F(u)\phi_x dxdt = -\int_{\Omega_+} [u_t - F(u)u_x]\phi dxdt + \int_C uv_t - F(u)v_x$$

$$= \int_C u_+ v_t - F(U_+)v_x dl$$

using the Greens' Theorem. Since $[u_t - F(u)u_x]$ vanishes almost everywhere on Ω_+, the integral equation is equivalent to the PDE. A similar argument may be applied to the region Ω_-. This implies that

$$[u_+ - u_-]v_t - [F(u_+) - F(u_-)]v_x = 0$$

It is well known that for a given curve, $x = s(t)$, the tangent vector is given by $\tau = (\dot{s}, 1)$, so that the normal vector is $(v_x, v_t) = -1, \dot{s})$. This leads to the relation

$$[u_+ - u_-]\dot{s} = [F(u_+) - F(u_-)]$$

which is nothing but the jump condition. An important remark about the weak solutions is that they are not always unique.

7.7 Waves in a Polytropic Gas

In the case of a polytropic gas, considering w which is the heat function, one of the RH conditions can be simplified to yield

$$v_1 - v_2 = [(p_2 - p_1)(V_1 - V_2)^2] \tag{7.63}$$

which can be simplified to yield

$$w_1 - w_2 + (1/2)(V_1 + V_2)(p_2 + p_1) = 0 \tag{7.64}$$

Replacing the heat function w by $\epsilon + pV$, where ϵ is the internal energy, the above relation can be simplified to yield

$$\epsilon_1 - \epsilon_2 + (1/2)(V_1 - V_2)(p_1 + p_2) = 0 \tag{7.65}$$

© 2015 by Taylor & Francis Group, LLC

For a given p_1, V_1, equation (7.64) or (7.65) gives the relation between p_2 and V_2. This relation in the literature is often referred to as the shock adiabatic or the Hugoniot adiabatic.

The heat function may be simplified to yield

$$w = c_p T = pV/(\gamma - 1) = c^2/(\gamma - 1) \tag{7.66}$$

Substituting the above relation into equation (7.64), we can determine the quantities p_1, V_1, p_2, etc. The ratio V_2/V_1 can be shown to be a monotonically decreasing function of the ratio p_2/p_1, tending to the finite limit of

$$\frac{V_2}{V_1} = (\gamma - 1)/(\gamma + 1) \tag{7.67}$$

The curve showing p_2 as a function of V_2 for a given value of p_1 and V_1 is given in Figure 7.5. The curve is a rectangular hyperbola with asymptotes at $V_2/V_1 = (\gamma - 1)/(\gamma + 1)$, $p_2/p_1 = -(\gamma - 1)/(\gamma + 1)$. Only the upper part of the curve is shown in Figure 7.5.

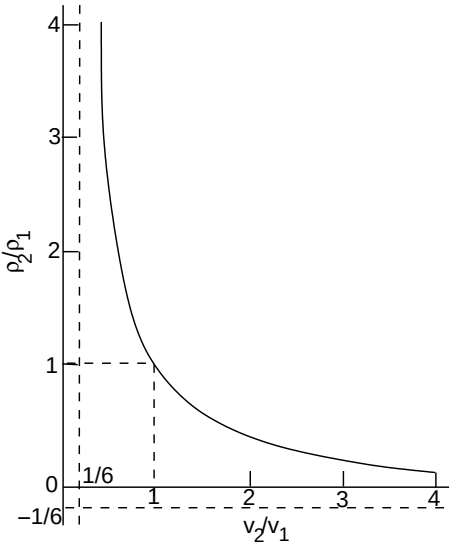


FIGURE 7.5: Ratio of the pressure as a function of V_2/V_1.

From the ratio of the temperatures on the two sides of the discontinuity we find that

$$\frac{T_2}{T_1} = \frac{\dfrac{p_2}{p_1}(\gamma + 1)p_1 + (\gamma - 1)p_2}{(\gamma - 1)p_1 + (\gamma + 1)p_2} \tag{7.68}$$

The mass flux can be easily calculated as

$$j^2 = [(\gamma - 1)p_1 + (\gamma + 1)p_2]/2V_1 \tag{7.69}$$

© 2015 by Taylor & Francis Group, LLC

the velocities of the propagation of the shock wave relative to the gas before and behind it are given by

$$v_1^2 = (1/2)V_1[(\gamma-1)p_1 + (\gamma+1)p_2]$$
$$= (1/2)\frac{c_1^2}{\gamma}\left[\gamma - 1 + (\gamma+1)\frac{p_2}{p_1}\right]$$
$$v_2^2 = (1/2)V_1\frac{[(\gamma-1)p_1 + (\gamma-1)p_2]^2}{(\gamma-1)p_1 + (\gamma+1)p_2}$$
$$= (1/2)\frac{c_2^2}{\gamma}[\gamma - 1 + (\gamma+1)\frac{p_1}{p_2}$$

Their difference is calculated as

$$v_1 - v_2 = (2V_1)(p_2 - p_1)(1/2)/(\gamma-1)p_1 + (\gamma+1)p_2 \tag{7.70}$$

The ratios of the densities, pressure, and temperature may be calculated to yield the following:

$$\rho_2/\rho_1 = v_1/v_2 = (\gamma+1)M_1^2/[(\gamma-1)M_1^2 + 2] \tag{7.71}$$
$$p_2/p_1 = 2\gamma M_1^2/(\gamma+1) - (\gamma-1)/(\gamma+1) \tag{7.72}$$
$$T_2/T_1 = [2\gamma M_1^2 - (\gamma-1)][(\gamma-1)M_1^2 + 2]/(\gamma+1)^2 M_1 \tag{7.73}$$

The expression for the Mach number M_2 in terms of M_1 is given by

$$M_2^2 = \frac{2 + (\gamma-1)M_1^2}{2\gamma M_1^2 - (\gamma-1)} \tag{7.74}$$

Simplifying the above expression yields

$$2\gamma M_1^2 M_2^2 - (\gamma-1)(M_1^2 + M_2^2) = 2 \tag{7.75}$$

For strong shocks wherein $(\gamma-1)p_2$ is very large as compared to $(\gamma+1)p_1$, we have the following results:

$$\frac{V_2}{V_1} = \frac{\rho_1}{\rho_2} = \frac{\gamma-1}{\gamma+1} \tag{7.76}$$
$$\frac{T_2}{T_1} = \frac{(\gamma-1)p_2}{(\gamma+1)p_1} \tag{7.77}$$

The ratio T_2/T_1 tends to infinity with p_2/p_1. The temperature discontinuity turns into a shock wave, as does the pressure discontinuity. This discontinuity, in principle, can be large. The density ratio, however, tends to a constant limit. In the case of a monoatomic gas, the limit is given by $\rho_2 + 4\rho_1$, and for a diatomic gas, it is $\rho_2 + 6\rho_1$. The corresponding velocities of propagation for the strong shock wave are given by

$$v_1 = [(1/2)(\gamma+1)p_2V_1]^{1/2}, \qquad v_2 = \left[\frac{(1/2)(\gamma-1)^2 p_2 V_1}{\gamma+1}\right]^{1/2} \tag{7.78}$$

© 2015 by Taylor & Francis Group, LLC

The velocities increase as the square root of the pressure. As mentioned in the previous section on weak shocks, let us define a quantity $z = (p_2 - p_1)/p_1$ and expand in terms of this quantity. In terms of z, the expression for the density ratio and sound speed can be written as

$$M_1 - 1 = 1 - M_2 = (\gamma + 1)z/4\gamma \quad (7.79)$$

$$c_2/c_1 = 1 + (\gamma - 1)z/2\gamma \quad (7.80)$$

$$\rho_2/\rho_1 = 1 + z/\gamma - (\gamma - 1)z^2/2\gamma \quad (7.81)$$

7.8 An Application of Shock Waves in the Sun

Solar flares and coronal mass ejections (CMEs) are explosive phenomena in the solar atmosphere, viable of launching global large-amplitude coronal disturbances and shock waves. The longest-known signatures of coronal shock waves are radio type II bursts and Moreton waves ([59], [60]). A type II burst is excited at the local plasma frequency (and/or harmonic) by a fast-mode MHD shock. As the shock propagates outwards through the corona, the emission drifts slowly (in comparison with fast drift type III emission) towards lower frequencies due to decreasing ambient density. Radial velocities, inferred from the emission drift rates by using various coronal density models, are found to be on the order of 1000 kms^{-1}. The Moreton wave is a large-scale wave-like disturbance of the chromosphere, observed in Hα, which propagates out of the flare site at velocities also on the order of 1000 kms^{-1}. In this respect it is worth noting that first indications of global coronal disturbances were provided by flare-associated activations of distant filaments.

The MHD model unifying both phenomena in terms of the fast-mode shock wave was proposed by Uchida[61]. According to his sweeping-skirt scenario, the Moreton wave is the surface track of the fast-mode MHD coronal shock propagating out of the source region along valleys of low Alfven velocity, i.e., being refracted from the high Alfven velocity regions and enhanced in low velocity regions. At larger heights, the shock causes the type II burst. An EUV counterpart of the Moreton wave was reported by Neupert[62], and a decade later these coronal disturbances were directly imaged by the Extreme Ultraviolet Imaging Telescope on the Solar and Heliospheric Observatory (SoHO). The discovery of Moreton waves ([63], [171]) prompted a search for wave signatures in other spectral domains. Soon, the Moreton wave associated disturbances were revealed in soft X-rays, He I 10830 A° and microwaves. It is important to note that some of propagating EUV signatures denoted as are probably not wave phenomena, but rather a consequence of some other processes related to the large scale magnetic field reconfiguration. Several disturbances of this kind are usually much slower and more diffuse than those representing the

© 2015 by Taylor & Francis Group, LLC

coronal counterpart of the Moreton waves and those accompanied by type II bursts. Such nonwave events either could be a consequence of the CME associated field line opening or could be caused by various forms of coronal restructuring driven by the eruption. Generally, large-amplitude MHD waves in the corona are closely associated with CMEs and flares. The source of the coronal wave seems clear in events where the CME is accompanied only by a very weak/gradual flare-like energy release. In such cases the type II bursts characteristically start in the frequency range well below 100 MHz.

The wave fronts are seen in emission in the center and the blue wing of the H_α line, whereas in the red wing they appear in absorption. This is interpreted as a compression and subsequent relaxation of the chromosphere, due to the increased pressure behind the coronal shock sweeping over the chromosphere[61]. Such a behavior strongly favors the interpretation in terms of the freely propagating large-amplitude simple-wave. Further supporting evidence for such an interpretation is found in the deceleration of the wavefront, elongation of the perturbation profile, and decreasing amplitude of the disturbance. In this respect it should be emphasized that there are two ways to form a simple wave shock pattern. The straightforward option is the formation of the shock by a temporary 3D piston effect, which can be caused either by the flare-volume expansion or by the initial lateral expansion of the CME. On the other hand, distant flanks of a bow shock also have simple-wave characteristics. This suggests that flares which occur in the core of the active region and do not have remote extensions towards quiet regions away from sunspots, are not likely to cause a Moreton wave.

The Moreton wavefront usually becomes detectable at distances on the order of 100 Mm from the source region, most often becoming clearly recognizable in the range of 100 to 150 Mm. As in the case of high-frequency type II bursts, the Moreton wave onset is closely associated with the flare impulsive phase, usually being delayed by a few minutes. Thus, the onset of the type II burst and the Moreton wave appearance are closely linked, implying that the Moreton wave becomes prominent only after the shock has been formed. Such a short time/distance for shock formation requires an extremely impulsive acceleration of the source region. Since the source region expansion has to be accelerated to a velocity on the order of 1000 kms^{-1} within a minute or so, this requirement favors the flare scenario, since flares typically develop on a shorter time scale than CMEs.

Moreton waves are generally observed to be closely associated with the flare impulsive phase, which often coincides also with the acceleration phase of the associated CME. Moreton waves are observed to propagate perpendicular to the magnetic field, and the initial magneto sonic Mach numbers are estimated to lie in the range of $\mathrm{M_{ms}} \approx 1.4 - 4$, suggesting that they are at least initially shocked fast-mode waves. These results indicate that Moreton waves are a consequence of shocks formed from large amplitude waves that decay to ordinary fast magneto sonic waves, which is in line with the flare-initiated blast wave scenario. Further evidence for the close association with

© 2015 by Taylor & Francis Group, LLC

shocks is the quasi simultaneous appearance of Moreton waves and radio type II bursts, which are one of the best indicators of coronal shocks. Today the relation of waves with Moreton waves and their generation mechanism are very much debated.

7.8.1 An Example from Earth's Bow Shock

If we look at the mean free path for the collision of particles in the solar wind with the boundary of the earth's magnetosphere, it is of the order of 10^8 km, which implies that the Coulomb collisions may be neglected completely. This also implies that the collisonal viscosity μ is negligible, which will make it the observation of shock waves difficult as there is no mechanism which will prevent wave steepening. However, from the satellite observations, we have strong evidence of a well-defined shock transition, whose thickness is typically on the order of 100 km. This shock in literature is referred to as the "bow shock," which lies between the solar wind and the earth's magnetosphere. A typical observation of such an event is presented in Figure 7.6.

We enumerate some fundamental differences between ordinary, collision-dominated and collisionless shocks. The important aspect is that the plasma is in general not in thermodynamic equilibrium behind the shock. Depending on the dissipation mechanism, the electron and ion pressure may be different from each other. The jump conditions may not be able to determine the downstream state of the plasma. The thickness of the collision-dominated shock is only a few collision mean free paths. However, for a collisionless shock, the thickness over which the shock occurs may be much larger. It is difficult to define a shock thickness uniquely, as there are more than one scale lengths in the system, which depends on the dissipation mechanisms considered.

A discussion (theoretical) on the nature of shock waves in collisionless plasmas from the dispersive and dissipative point of view is presented in the following section.

7.9 Shock Waves in Collisionless Plasmas

It is well known from gas dynamics that heating of a medium by a strong shock wave is very efficient, and it seems natural to apply shock heating to plasmas, as well; in the framework of nuclear fusion research, for example, heating of a plasma to high temperatures in a short time is one of the most important problems. However, classical shock waves where the dissipation is provided by particle collisions can only be of limited importance. In the early phase of a 6-pinch implosion, a collision-dominated shock could be produced. As the plasma temperature rises, binary collisions become rare and the mean free path soon exceeds the dimensions of the plasma vessel. So if shock waves

© 2015 by Taylor & Francis Group, LLC

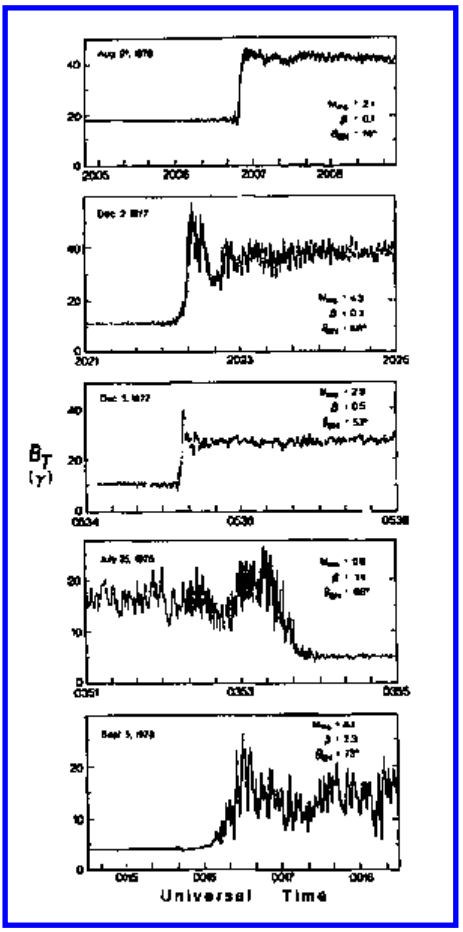


FIGURE 7.6: An example of collisionless shock produced by the interaction of the solar wind and the Earth's magnetosphere, from Russel et al.[19].

are possible in a high-temperature, i.e. nearly collisionless plasma, the dissipation must be due to some process other than binary collisions. Such processes do indeed exist; a collisionless plasma sufficiently far away from thermal equilibrium usually responds by becoming unstable and producing some kind of turbulence which gives rise to effective or "anomalous" transport coefficients. The first conclusive evidence of the existence of collisionless shocks came from astrophysics. It had been known for quite some time that the plasma flow from the Sun, the solar wind, is highly supersonic when encountering the magnetic field of the Earth, and a shock wave should be formed. However, satellite observations, nevertheless, clearly showed the presence of a well-defined shock transition with a thickness of the order of 100 km, i.e., a distance over which Coulomb collisions are completely negligible.

© 2015 by Taylor & Francis Group, LLC

In laboratory experiments, first evidence of collisionless shock waves was reported in 1965, and research on this subject has since developed into a discipline of its own. Most of the experiments are performed in 6- or z-pinches. A fast-rising magnetic field at the outer boundary of the plasma acts as a magnetic piston which generates a shock wave propagating towards the axis. Many experimental difficulties appear if one wants to have a shock neatly detached from the piston field, and there are the problems entailed in achieving good diagnostics. In the theoretical investigations of collisionless shock waves there are essentially two different aspects that have been considered so far. The first refers to laminar shock waves, where dispersion limits nonlinear steepening and a (trailing or leading) wave train is generated. Dissipation enters only by including some process, collisional or collisionless, leading to damping of the wave train. For small Mach numbers the theory of laminar shock waves is fairly complete. However, direct practical applications are limited. Even for rather weak shocks, experimentally observed profiles are often quite different from the dispersive wave trains, being rather monotonic instead of oscillatory, thus indicating the presence of efficient collisionless dissipation processes. The separate investigation of these processes such as the current-driven ion-sound instability responsible for the anomalous resistivity represents the second aspect of collisionless shock theory. However, to derive theoretical expressions for anomalous transport coefficients, a knowledge of the nonlinear behavior of an instability is required, which is a very complicated theoretical problem.

In the last few years, understanding of collisionless shock wave phenomena has been substantially improved both by very elaborate experimental investigations and by numerical computations. Among the experiments there are, for example, detailed measurements of the frequency and wavenumber spectra in so-called resistive shock waves and the high-resolution plasma and field measurements in the Earth's bow shock, which is a high-Mach-number shock wave. On the theoretical side, in addition to analytical methods, extensive numerical simulations, which have recently become possible, have proved to be a valuable tool for investigating collisionless plasma processes. They have improved our understanding of currentdriven turbulence producing anomalous resistivity and of anomalous ion heating in high-Mach-number shocks. Although much work remains to be done, it is fair to say that, owing to the recent progress in experiments and theory, the picture of collisionless shock wave phenomena has become much clearer.

In the theory of shock waves, an interesting question is how to determine the properties of the medium after passage of the shock in terms of the unperturbed quantities such as density, pressure, and Mach number. Assuming that the shock profile is (quasi-) stationary in a certain frame of reference, a number of relations between upstream and downstream quantities are given by the jump conditions, which derive directly from the conservation of mass, momentum, and energy flux and, perhaps, further relations such as magnetic flux conservation in a magneto hydrodynamic (MHD) system. In usual gas dynamics and MHD-theory, these jump conditions are sufficient to determine

© 2015 by Taylor & Francis Group, LLC

the downstream state although a theory of the shock structure and the dissipation processes leading to entropy production requires a kinetic description, especially at high Mach numbers. In the case of a collisionless shock wave, the plasma is, in general, not in local thermodynamic equilibrium behind the shock transition. For example, electron and ion pressures may be quite different from each other, depending on the dissipation mechanism. Here the jump conditions do not completely determine the downstream state, a description of the shock structure itself being needed. This would, e.g., provide information on the ion or electron thermalization length, knowledge of which is useful in planning the dimensions of a plasma system to be heated by a shock.

7.9.1 Dispersive Shock Waves

We briefly discuss the properties of dispersive shock waves. A small amount of dissipation produces a damped wave train, the final state of which is usually identified with the bottom of the well of the corresponding nonlinear oscillator problem. A brief discussion of the nature of collisionless dissipation processes follows. Recent developments, numerical simulation, and very detailed experimental measurements have considerably clarified the situation, which was in danger of being suffocated under the number of different instabilities and nonlinear saturation processes.

Let us illustrate the subtle difference between gas dynamic and weak shocks in a collisionless plasma. To begin with, we shall mention some well known salient features of shock formation in a gas. Consider a plane wave front with sufficiently small amplitude (linear theory). Such a wave front can be described by the wave equation as

$$\frac{\partial u}{\partial t} + (u \pm c_s)\frac{\partial u}{\partial x} = \mu \frac{\partial^2 u}{\partial x^2} \tag{7.82}$$

where u is the amplitude of the velocity and c_s is the velocity of sound. We shall assume that $u \ll c_s$, μ is the kinematic viscosity ($\mu \approx \lambda_c c_s$, where λ_c is the mean free path). Let us deal with the development of an initial perturbation $u_0(x)$. If we also assume that $\lambda_c \ll Lu/c_s$, where L is the scale of the disturbance, one can in principle neglect the right-hand side of the above equation. The general solution of the above equation, neglecting the right-hand side can be written simply as

$$u = u_0(x - t(u \pm c_s)) \tag{7.83}$$

The above expression implies that any initial pulse will steepen until

$$\partial u/\partial x = u_0'(1 + u_0')^{-1}$$

becomes singular at some point. The steepening will be delayed or suppressed if the dissipation term is included as in the initial wave equation. The linearized solution implies that the pulse propagates with the same velocity c_s. In the

© 2015 by Taylor & Francis Group, LLC

case of finite amplitude waves, the introduction of a nonlinear term $u\partial u/\partial x$ would give rise to steepening when parts of u are large enough to overtake slower one. This can be balanced by the dissipation term $\mu\partial^2 u/\partial x^2$.

In an ordinary Boltzmann gas, only acoustic modes exist, which are nondispersive, while a rarefied plasma can support many kinds of waves with different dispersion relations. For example, we have the magneto acoustic modes and the ion-acoustic modes, to name two.

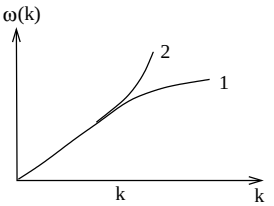


FIGURE 7.7: Types of dispersion curves.

In the case of a cold plasma, magneto acoustic waves propagating perpendicular to the magnetic field have a dispersion curve of type 1 as shown in Figure 7.7., while for obliquely propagating waves, the dispersion is of type 2. This is typically the Whistler branch, whose relation is given by $\omega^2 \approx k^2 v_A^2(1 + k^2\cos^2\theta/\omega_{pl}^2$ and θ is the angle between the wave vector and the magnetic field.

For sufficiently small amplitudes, a dispersive system can be described by the Korteweg de Vries equation (KdV) as

$$\frac{\partial u}{\partial t} + (u \pm c_s)\frac{\partial u}{\partial x} = a\frac{\partial^3 u}{\partial x^3} \tag{7.84}$$

This equation is similar to the equation we have mentioned above, with the right-hand side representing dispersion rather than dissipation. The linear dispersion relation for the above equation can be written down as

$$\omega = \pm kc_s + ak^3$$

Let us discuss the essential features of dispersive shock waves for the above equation. Consider solutions of the type $u = u(x+Mc_st)$, a wave with constant velocity Mc_s propagating, in the negative direction. Here M is the Mach number and it is assumed that $M - 1 \ll 1$. The KdV equation can be transformed to a nonlinear oscillation problem with the transformation

$$V(u) = -\frac{1}{6a}(u - c_s)^3 + b(u - c_s) \tag{7.85}$$

Here a is the dispersion parameter and b is the integration constant and the energy of the oscillator. The different solutions which depend on a and b are determined. We seek finite amplitude solutions so that $|u| < \infty$ everywhere.

© 2015 by Taylor & Francis Group, LLC

This implies that $ab > 0$. The different solutions are either periodic nonlinear waves or symmetric solitary waves (solitons) with $u = u_0 = Mc_s$, for $x \to \pm\infty$ with a maximum amplitude of u_m. It can be shown that for $a > 0$, there are no shock-like solutions. However, an oscillatory shock wave is found if a small amount of dissipation is introduced into the KdV equation. Starting at u_0, the oscillator will lose energy and instead of returning to this value, it will oscillate down to the potential minimum u_1. For $a > 0$, a shock with a trailing wave train is generated, whereas for $a < 0$, there is a shock with a leading wave train, a precursor to the potential minimum u_1, being the upstream state as compared to u_0 which is the downstream state (see Figure 7.8).

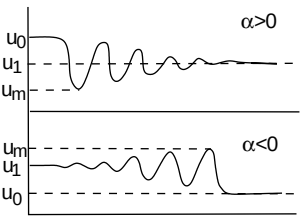


FIGURE 7.8: Dispersive shocks with trailing ($a > 0$) and leading ($a < 0$) wave train.

Typically, in a gas dynamic shock theory of a one-component gas, the transition layer is characterized by a single parameter, namely, the width of the shock, $\triangle$. However, in collisionless plasmas, such a simple description is no longer possible, considering the fact that it exhibits more complicated shock structures. In a simple, dispersive shock, sometimes, there are more than one scale, the dispersion length, and the dissipation scale. When the Mach number is large, several dissipation processes appear which lead to more complicated shock structures.

For large (or finite) amplitudes, one has to replace equation (7.82) by an equation (or a set of equations) which allows the nonlinear aspects. In the limit of vanishing dissipation, one can reduce the steady state equation to an equation of a nonlinear oscillator of the form (7.82). The subtle difference between the shock and nonlinear oscillator is that the former represents a closed system, where any oscillatory energy dissipated must increase the internal plasma energy, while in the latter case, the quasi-potential, being a function of the pressure, changes during the wave decay, however small the rate of dissipation may be. We shall illustrate this point by considering the stationary solutions of the dissipation free system described by the following

© 2015 by Taylor & Francis Group, LLC

set of equations:

$$\frac{d}{dx}(nu) = 0 \tag{7.86}$$

$$\frac{d}{dx}\left(m_i n u^2 + p + \frac{B^2}{8\pi}\right) = 0 \tag{7.87}$$

$$\frac{d}{dx}(uj) = \frac{n_e^2}{m_e}\left(E_y - \frac{u}{c}B\right) \tag{7.88}$$

$$\frac{dB}{dx} = -\frac{4\pi}{c}j \tag{7.89}$$

$$\frac{d}{dx}\left(\frac{P}{n^2}\right) = 0 \tag{7.90}$$

We assume that the wave propagation is in the x direction, while the magnetic field B is in the z direction. The plasma velocity u is also in the x direction. j is the current density and $m_e \ll m_i$. The two Mach numbers are defined as

$$M_A^2 = 4\pi n_0 m_i u_0^2/B_0^2, \qquad M_s^2 = n_0 m_i u_0^2/2p_0$$

Equations (7.86−7.90) can be transformed to an equation of the form

$$\frac{d^2B}{d\tau^2} = Bu(B) - 1 \equiv -\frac{dV(B)}{dB} \tag{7.91}$$

where $d\tau = dx/u$, and $u(B)$ is the solution of the equation

$$u + \frac{1}{2u^2M_s^2} = 1 + \frac{1}{2M_0^2} - \frac{1}{2M_A^2}B^2 \tag{7.92}$$

with

$$\frac{1}{M_0^2} = \frac{1}{M_A^2} + \frac{1}{M_s^2}$$

By simple algebra one can show that the potential well is located at

$$B_b = (1/2)[(1 + 8M_0^2)^{1/2} - 1] \tag{7.93}$$

The above expression would be the final value of B in the downstream state if the potential V did not undergo any change during the dissipation process. The jump conditions are written as

$$[nu] = 0 \tag{7.94}$$

$$\left[nm_i u^2 + p + \frac{B^2}{8\pi}\right] = 0 \tag{7.95}$$

$$[uB] = 0 \tag{7.96}$$

$$\left[(1/2)nm_i u^3 + 2u\left(p + \frac{B^2}{8\pi}\right)\right] = 0 \tag{7.97}$$

© 2015 by Taylor & Francis Group, LLC

$[\cdots]$ denotes the difference between the upstream and downstream values. The final value of B can be calculated as

$$B_f = \frac{3M_0^2}{2 + M_0^2} \tag{7.98}$$

It is easy to realize that $B_b - B_f$ becomes arbitrary large for larger values of M_0, i.e., the magnetic energy increase across the shock is smaller in the real system, while the increase in the pressure is large.

In what follows, we shall study the behavior of the collisionless plasma, when the dissipation is introduced. In the case of gas dynamics, the increase in the entropy across the shock is determined from the jump condition

$$[s] = c_c \left[\ln \frac{p}{n^\gamma}\right] \tag{7.99}$$

The stationary solution of the Vlasov equation has the property that the entropy flux is conserved. In terms of a relation, we have

$$\frac{\partial}{\partial \mathbf{x}} \cdot \int d^3 v \mathbf{v} \mathbf{f} \ln \mathbf{f} = \mathbf{0} \tag{7.100}$$

In general the plasma state has to be time dependent as the shock transition is stationary only in the sense that it is an average over the rapid motions. The entropy may be defined as the average distribution function $\langle f \rangle$. The divergence of the entropy flux can be written as

$$\frac{\partial}{\partial \mathbf{x}} \cdot \int d^3 v \mathbf{v} \langle \mathbf{f} \rangle \ln \langle \mathbf{f} \rangle = \frac{\mathbf{e}}{\mathbf{m}} \int \mathbf{d^3 v} \langle (\delta \mathbf{E} + \frac{\mathbf{v}}{\mathbf{c}} \times \delta \mathbf{B}) \delta \mathbf{f} \rangle \times \frac{\partial \ln \langle \mathbf{f} \rangle}{\partial \mathbf{v}} \tag{7.101}$$

where

$$f = \langle f \rangle + \delta f, \qquad \mathbf{E} = \langle \mathbf{E} \rangle + \delta \mathbf{E}$$

The transition from a laminar to a turbulent shock structure is conveniently described by the development of an instability. Instabilities are usually classified as macroscopic or microscopic, and, within the latter category, a distinction is made between resonant and nonresonant. Resonant instabilities directly give rise to collisionless dissipation. They are driven by the interaction (resonance) of a small fraction of particles with a wave, which in the presence of a broad spectrum of waves leads to turbulence in phase space, i.e., to collisionless dissipation. More details on shock waves in collisionless plasmas can be obtained from Biskamp[64].

7.10 Shocks in MHD

Compressibility effects become important when the flow speed approaches the velocity of sound lead to shock waves. These waves involve nonlinear processes and they are formed in space and produced by collisionless plasmas.

© 2015 by Taylor & Francis Group, LLC

There are enough reasons to believe, both with theoretical and observational results, that collisionless plasmas do produce shock waves. In the previous section on collisional plasmas, we gave some justification for the formation of shocks as well as provided observational evidence for bow shocks due to interaction of the solar wind with supersonic fluid flows. It is possible that space plasmas support both steady-state as well as transient shocks depending on the physical mechanisms involved. As mentioned earlier, the nature of shocks in both collisional and collisionless plasmas depends a lot on the dissipation mechanisms of the system.

In what follows, we shall discuss some basic concepts and definitions in MHD before a detailed analysis of shocks present in MHD. Consider a fluid moving with a uniform velocity **U**. For a compressible fluid, this perturbation will propagate in the form of a compressional sound wave. Let c_s be the local sound velocity and **n** a unit vector that defines the direction of propagation. The velocity of the wave, relative to a fixed point O, is

$$\mathbf{V} = \mathbf{U} + c_s \mathbf{n}$$

The magnitude of **V** depends on the direction of propagation and is maximum when c_s and **U** are in the same direction and minimum when they are opposite each other. The wave propagation depends on the ratio of the the sound speed and the flow velocity. If we denote this ratio in terms of a cone with an angle α, then

$$\sin \alpha = c_s / U$$

This angle is called the Mach angle and the ratio the Mach number. Depending on whether $M < 1$ or $M \geq 1$, the flow is termed subsonic or supersonic, respectively. The behavior of the shock waves depends on the steepening of the compressional waves. The wave tends to steepen and changes its shape as it propagates, until the flow becomes nonadiabatic. In such a scenario, viscous and heat conduction effects become important and a shock wave develops in the fluid until a balance between dissipation and steepening is achieved. A parameter that is used to categorize shocks is the angle between the magnetic field direction and the shock normal θ_{Bn}. When $\theta_{Bn} = 90°$, then it is a perpendicular shock, while for $\theta_{Bn} = 0°$, it is a parallel shock. Angles between 0 and 90 are termed oblique shocks. A specific case when the angles lie between $0 < \theta_{Bn} < 45$ and $45 < \theta_{Bn} < 90$ are called quasi parallel and quasi perpendicular shocks, respectively.

In MHD, shocks are designated super critical and subcritical depending on whether the Mach number is greater than or less than 2 to 3. An illustration of a bow shock showing the regions where parallel and perpendicular shocks are formed is given in Figure 7.9.

MHD shocks can be approximated to shocks in ordinary fluids using the conservation equations, with the exception that the normal component of the flow velocity $U_n \neq 0$ and other factors which will be discussed below.

© 2015 by Taylor & Francis Group, LLC

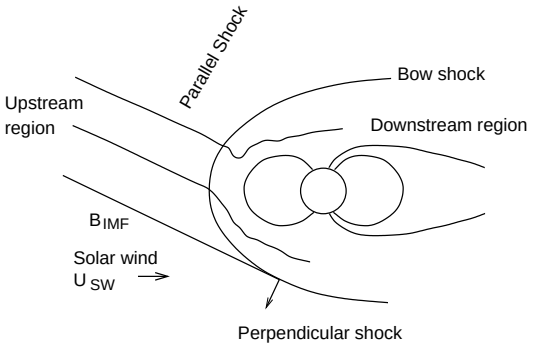


FIGURE 7.9: Schematic view of a bow shock − The formation of parallel and perpendicular shocks is also illustrated, from Parks[151].

7.10.1 Parallel Shocks

Parallel shocks are formed when the geometry of the flow with the magnetic field $\mathbf{B}$ is parallel to the direction of the shock normal $\mathbf{n}$. This implies that $B_t = 0$ and $B = B_n$. The tangential component is not present. The continuity equations derived for ordinary shocks also apply for MHD. In ordinary shocks, the collisions dominate and thus one can assume that the energy satisfies the equipartition condition in all directions, which means $w_\perp = 2w_\parallel$. Thus, the use of a Maxwellian distribution is justified which implies an adiabatic equation of state. The relation between the internal energy and the ratio of specific heats is written as

$$w_\perp = \left(\frac{\gamma - 1}{\gamma}\right) \frac{p_\perp}{\rho} \tag{7.102}$$

In this case, the magnetic pressure is unimportant and so only the gas pressure is included. For parallel shocks, we can introduce a transformation to a shock frame of reference where $U_t = 0$. The velocity will be parallel to the magnetic field. We can write

$$\mathbf{B} = \alpha \rho \mathbf{U} \tag{7.103}$$

With $\mathbf{U} \| \mathbf{B}$, α is a constant along the streamline. Since the discontinuity in the magnetic field $[B] = 0$ and $[\rho U] = 0$, it implies that $[\alpha] = 0$. In the case of parallel shocks, its passage will not affect the structure of the magnetic field. In what follows, we will see that such a simple scenario is not possible for perpendicular shocks.

7.10.2 Perpendicular Shocks

Let us define a reference frame in which the velocity of the shock is $\mathbf{V}$ and the flow velocity $\mathbf{U}$. Let us also use the terminology laboratory system when $\mathbf{U} = \mathbf{0}$ and shock system when $\mathbf{V} = \mathbf{0}$. In general it is convenient to use the shock system for discussing the conservation laws.

© 2015 by Taylor & Francis Group, LLC

Choose the magnetic field and electric field as $\mathbf{B} = (0, 0, B)$ and $\mathbf{E} = (0, E, 0)$ so that the shock is confined to the yz plane. Also assume that $\mathbf{U} = (U, 0, 0)$ and N is the number density . The particles tend to move along the magnetic field so that the normal component of U is related to the electric field tangent to the shock. For the geometry under consideration, Maxwell's equation can be simplified as

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t} \tag{7.104}$$

Integrating the above equation across the shock leads to

$$E_2 - E_1 = -\frac{\partial \Phi}{\partial t} \tag{7.105}$$

where $\Phi = \int B dx$ is the magnetic flux. If we assume that the shock structure is infinitely thin, then $\Phi = 0$ in the shock. It is important to impose the condition that the electric field be the same on both sides, $E_2 \approx E_1 = E$.

The fluid velocities on the upstream and downstream can be written as

$$\begin{aligned} U_1 &= \frac{E}{B_1} \\ U_2 &= \frac{E}{B_2} \end{aligned} \tag{7.106}$$

which implies that

$$\frac{U_1}{U_2} = \frac{B_2}{B_1} \tag{7.107}$$

Using conservation of particles, a simple algebra leads to

$$\frac{U_1}{U_2} = \frac{N_2}{N_1} \tag{7.108}$$

For $U_1 > U_2$, the above relations imply that the density of particles and magnetic fields are compressed by the same amount across the shock. For a difference in the magnetic field across the shock, there should be a current present. If we assume that $B_2 > B_1$, the Maxwell's equation simplifies to

$$\frac{\partial B}{\partial x} = -\mu_0 J_y \tag{7.109}$$

As in the previous case for electric fields, let us integrate the above relation to get

$$B_2 - B_1 = \mu_0 \int J_y dx \tag{7.110}$$

J_y is the result of magnetization and drift currents, which in turn is due to gradients in the magnetic fields across the shock. For a shock geometry which

© 2015 by Taylor & Francis Group, LLC

is curved, a curvature current needs to be included to the gradient current. The conservation of momentum equation requires that

$$\frac{\partial}{\partial x_j}(\pi_{ij} + T_{ij}) = 0 \tag{7.111}$$

where π_{ij} and T_{ij} are the momentum transfer and electromagnetic stress tensors, respectively. For the plane geometry, integrating the above relation across the shock yields

$$(\pi_{ix} + T_{ix})_1 = (\pi_{ix} + T_{ix})_2 \tag{7.112}$$

The electromagnetic stress tensor for a plane geometry is a diagonal tensor. The particle momentum stress tensor is given by

$$(\pi_{xx})_1 = N_1 m U_1^2 + (p_{xx})_1, \qquad (\pi_{xx})_2 = N_2 m U_2^2 + (p_{xx})_2 \tag{7.113}$$

Defining the total pressure as $p^* = p + B^2/2\mu_0$, we can show that

$$\rho_1 U_1^2 + p_1^* = \rho_2 U_2^2 + p_2^* \tag{7.114}$$

where $\rho = Nm$ is the mass density. This expression is similar to the momentum conservation in ordinary shocks, except that the pressure is the total pressure which includes the magnetic pressure. Using the conservation of mass results in

$$\rho_1 U_1 (U_1 - U_2) = p_2^* - p_1^* \tag{7.115}$$

The next step is to examine the conservation of energy equation, which if steady, implies $\boldsymbol{\nabla}\cdot(\mathbf{q} + \mathbf{S}) = \mathbf{0}$. Here $\mathbf{q}$ and $\mathbf{S}$ are the energy flow and the Poynting flux, respectively.

Integration of equation (7.115) across a plane shock yields

$$(q_x + S_x)_1 = (q_x + S_x)_2 \tag{7.116}$$

The x component of $\mathbf{S}$ can be written as

$$S_x = \frac{(\mathbf{E}\times\mathbf{B})_\mathbf{x}}{\mu_0} = \frac{UB^2}{\mu_0}$$

q_i is defined by

$$q_i = \int \frac{mv^2}{2} v_i f(r, v) d^3 v \tag{7.117}$$

where $v_i = \langle v_i \rangle + \delta v_i$, $\langle v_i \rangle = u_i$, is the velocity averaged over the distribution function, and δv_i is the thermal velocity. The above equation can be simplified to yield

$$q_i = \frac{mn}{2}[u_i u^2 + u_i \langle \delta v^2 \rangle + 2u_i u_j \langle \delta v_j + u^2 \langle \delta v_i \rangle + \langle \delta v_i \delta v^2 \rangle + 2u_j \langle \delta v_i \delta v_j] \tag{7.118}$$

The different terms denote macroscopic internal energy carried by convection,

© 2015 by Taylor & Francis Group, LLC

thermal velocity, and heat conduction. If we ignore the heat conduction term, then the above equation reduces to

$$q_i = \frac{\rho_m}{2}(u^2 u_i + u_i\langle\delta v^2\rangle + 2u_j\langle\delta v_i \delta v_j\rangle) \tag{7.119}$$

Here $\rho_m = nm$. For a single fluid description, let $u_x = U$, so that the above equation can be simplified as

$$q_x = \frac{\rho_m}{2}(U^2 U + U\langle\delta v^2\rangle + 2U\langle\delta v_x \delta v_x\rangle) \tag{7.120}$$

The internal energy can be written as

$$\begin{aligned} \frac{m\langle\delta v^2\rangle}{2} &= w_\perp + w_\parallel \\ \frac{m\langle\delta v_x \delta v_x\rangle}{2} &= w_\perp \end{aligned} \tag{7.121}$$

where $w_\perp$ and $w_\parallel$ are, respectively, the internal energy associated with the perpendicular and parallel shocks. Substituting the above expression into q_x, we obtain

$$q_x = \frac{\rho U^2}{2}U + NU(2w_\perp + w_\parallel) \tag{7.122}$$

The energy equation can be simplified to yield

$$\rho_1 U_1^2 + N_1 U_1(2w_\perp + w_\parallel)_1 + \frac{B_1^2}{\mu_0}U_1 = \rho_2 U_2^2 + N_2 U_2(2w_\perp + w_\parallel)_2 + \frac{B_2^2}{\mu_0}U_2 \tag{7.123}$$

In order to determine the downstream conditions, given the upstream conditions, we also assume that the internal energy does not change across a shock, which means

$$w_{\parallel 1} = w_{\parallel 2} \tag{7.124}$$

With the above assumption and writing $p_\perp = Nw_\perp$ and $p_\parallel = Nw_\parallel$, the energy equation can be simplified to yield

$$\rho_1 U_1(U_1^2 - U_2^2) + 4(U_1 p_1^* - U_2 p_2^*) = 0 \tag{7.125}$$

Taking ratios by dividing by $p_1^* U_1$, yields

$$\left(\frac{p_2^*}{p_1^*} - 1\right)\left(1 + \frac{U_1}{U_2}\right) + 4\left(\frac{U_1}{U_2} - \frac{p_2^*}{p_1^*}\right) = 0 \tag{7.126}$$

Introducing the nondimensional variable $\xi = p_2^*/p_1^*$ as the ratio of the pressure on either side of the shock and the density, velocity, and magnetic field ratios as η, the above relation reduces to

$$(\xi - 1)(\eta + 1) + 4(\eta - \xi) = 0 \tag{7.127}$$

© 2015 by Taylor & Francis Group, LLC

η can be solved from the above equation to yield

$$\eta = \frac{(3\xi + 1)}{(\xi + 3)} \tag{7.128}$$

Equation (7.128) determines the compression ratio of the shock as a function of the pressure ratio in the upstream and downstream regions. For ξ large, the shock is termed a strong shock. In the limit $\xi \to \infty$, $\eta \to 3$. For ξ small, the shock is defined as a weak shock. An important observation from the above equation is the specific value of $\xi = 1$. In this case, there is no shock at all.

7.11 Nonlinear Studies

In this section, we shall briefly mention the Kortweg de Vries equation, which is a nonlinear equation arising in ion-acoustic waves. We have already mentioned the importance of the KdV equation in discussing the weak shock waves in collisionless plasmas. The assumptions are as follows: The ions are assumed to be cold and nondrifting, with respect to the electrons, i.e., ($T_i \ll T_e$). We shall use for simplicity a one-dimensional description. The inertial effects for the electron are neglected. The equation of state (isothermal) will be $P_e = n_e k_B T_e$. Being isothermal, T_e will be a constant. The momentum balance equation for the electrons is written as

$$m_e n_e \frac{dv_e}{dt} = 0 \tag{7.129}$$

$$e\frac{\partial \phi}{\partial x} = \frac{k_B T_e}{n_e}\frac{\partial n_e}{\partial x} \tag{7.130}$$

where $-e$ is the charge of the electron, $n_e(x,t)$ is the number density of the electron, and $\phi(x,t)$ is the electrostatic potential.

Equation (7.130) may be integrated to yield

$$n_e = n_0 \exp(e\phi/k_B T_e) \tag{7.131}$$

where n_0 is the background electron density which is assumed to be uniform.

The Poisson equation can be written as

$$\frac{\partial^2 \phi}{\partial x^2} = 4\pi(n_e - n_i) = 4\pi e(n_0 \exp(e\phi/k_B T_e) - n_i) \tag{7.132}$$

by replacing the expression for n_e as mentioned above. The equations of motion and continuity for the ions can be written as

$$\frac{\partial v_i}{\partial t} + v_i \frac{\partial v_i}{\partial x} = -\frac{e}{m_i}\frac{\partial \phi}{\partial x} \tag{7.133}$$

$$\frac{\partial n_i}{\partial t} + \frac{\partial n_i v_i}{\partial x} = 0 \tag{7.134}$$

© 2015 by Taylor & Francis Group, LLC

Here the number density is $n_i(x,t)$, the ion mean velocity is $v_i(x,t)$, and e and m_i, are the ion charge and mass, respectively.

Let us introduce certain nondimensional variables as

$$n = n_i/n_0. \quad u = u_i/c_{si}, \quad z = x/\lambda_D, \quad \tau = c_{si}t/\lambda_D, \quad \Phi = e\phi/k_BT_e$$

where $c_{si} = (k_BT_e/m_i)^{1/2}$ is the ion sound speed and $\lambda_D = (k_BT_e/4\pi n_0 e^2)^{1/2}$ is the Debye length.

The basic equations in nondimensional form can be reduced to

$$\frac{\partial u}{\partial \tau} + u\frac{\partial u}{\partial z} + \frac{\partial \Phi}{\partial z} = 0 \tag{7.135}$$

$$\frac{\partial n}{\partial \tau} + \frac{\partial nu}{\partial z} = 0 \tag{7.136}$$

$$\frac{\partial^2 \Phi}{\partial z^2} = e^{\Phi} - n \tag{7.137}$$

Equations (7.135−7.137) have a solution given by

$$\Phi = 3(M-1)sech^2\Big[(1/2(M-1))^{1/2}(z - M\tau)\Big] \tag{7.138}$$

M is the Mach number in the above expression. Introduce the change of variable $\epsilon = M - 1 \ll 1$. ϵ may be interpreted as a measure of the amplitude of the waves. The argument of the above equation may be written as

$$\epsilon^{1/2}(z - \tau) - \epsilon^{3/2}\tau \tag{7.139}$$

In deriving the KdV equation from equations (7.135−7.137), we introduce stretched variables

$$\chi = \epsilon^{1/2}(z - \tau) \tag{7.140}$$

$$\zeta = \epsilon^{3/2}\tau \tag{7.141}$$

The derivatives with respect to z and τ can be transformed as

$$\frac{\partial}{\partial z} = \epsilon^{1/2}\frac{\partial}{\partial \zeta} \tag{7.142}$$

$$\frac{\partial}{\partial \tau} = \epsilon^{3/2}\frac{\partial}{\partial \zeta} - \epsilon^{1/2}\frac{\partial}{\partial \chi} \tag{7.143}$$

We shall expand the dependent variables n, Φ, and u as a power series in ϵ as follows:

$$n = 1 + \epsilon n^{(1)} + \epsilon^2 n^{(2)} + \cdots \tag{7.144}$$

$$\Phi = \epsilon\Phi^{(1)} + \epsilon^2\Phi^{(2)} + \cdots \tag{7.145}$$

$$u = \epsilon u^{(1)} + \epsilon^2 u^{(2)} + \cdots \tag{7.146}$$

© 2015 by Taylor & Francis Group, LLC

The lowest order equations become (in ϵ)

$$\frac{\partial u^{(1)}}{\partial \chi} = \frac{\partial \Phi^{(1)}}{\partial \chi} \tag{7.147}$$

$$\frac{\partial u^{(1)}}{\partial \chi} = \frac{\partial n^{(1)}}{\partial \chi} \tag{7.148}$$

$$\Phi^{(1)} = n^{(1)} \tag{7.149}$$

which implies that $n^{(1)} = \Phi^{(1)} = u^{(1)}$, as all these have to vanish for $|\chi| \to \infty$.

The next higher order equations in ϵ can be written as

$$-\frac{\partial u^{(2)}}{\partial \chi} + \frac{\partial u^{(1)}}{\partial \chi} + u^{(1)}\frac{\partial u^{(1)}}{\partial \chi} + \frac{\partial \Phi^{(2)}}{\partial \chi} = 0 \tag{7.150}$$

$$-\frac{\partial n^{(2)}}{\partial \chi} + \frac{\partial n^{(1)}}{\partial \chi} + \frac{\partial (n^{(1)}u^{(1)}}{\partial \chi} + \frac{\partial v^{(2)}}{\partial \chi} = 0 \tag{7.151}$$

$$\frac{\partial^2 \Phi^{(1)}}{\partial \chi^2} - \Phi^{(2)} - \frac{(\Phi^{(1)})^2}{2} - n^{(2)} = 0 \tag{7.152}$$

A simple algebra by adding equations (7.150) and (7.151), using (7.152) to eliminate $\Phi^{(2)}$, and using the first-order result for eliminating $\Phi^{(1)}$ and $n^{(1)}$, we derive the equation

$$\frac{\partial u^{(1)}}{\partial \zeta} + u^{(1)}\frac{\partial u^{(1)}}{\partial \chi} + (1/2)\frac{\partial^3 u^{(1)}}{\partial \chi^3} \tag{7.153}$$

This is the famous Kortweg−de Vries equation. The dispersion parameter in this case is defined as $\xi = -(1/2)$.

7.11.1 Burger's Equation

In the last section of this chapter, we shall discuss various properties of Burger's equation, the Cole−Hopf transformation, the shock solutions, the N wave solution, etc. Wave phenomena, in general, are governed by nonlinear systems of partial differential equations, subject to certain physically motivated initial and/or boundary conditions. For example, the Navier−Stokes equations represent a typical example for such a system. In most of the cases, analytical solutions for these systems are far from being amenable and easy. However, using techniques such as the perturbation methods, the systems retain the essentials of the physical requirements. One of the best known nonlinear equations arising in physical systems, the Kortweg−de Vries equation, of which we have mentioned in this chapter as well as in previous chapters. Another interesting nonlinear equation which has been used as a model in physical problems, such as turbulence, traffic flow problems, and gas dynamics has interesting properties. We shall skip the detailed derivation of this

© 2015 by Taylor & Francis Group, LLC

equation for brevity. Interested readers can get a good idea about this equation from the book by Sachdev[159]. Burger's equation is given by

$$u_t + uu_x = \frac{\delta}{2}u_{xx} \tag{7.154}$$

Here t and x are the independent variables, u may be interpreted as velocity of a physical system, and δ is a parameter which may or may not be a constant. The suffix denotes partial derivative with respect to the independent variable.

Historically, the above equation was first derived by Bateman[65], who found some steady solutions, describing viscous flows. However, in subsequent years Burger's[66] used this equation to describe mathematical models of turbulence. In the context of gas dynamics, Hopf[67] and Cole[68] discussed this equation, by invoking a transformation (nonlinear), which in literature came to be called the Cole−Hopf transformation. The beauty of this transformation is that with this, the nonlinear Burger's equation is reduced to a simple heat equation. It is given by

$$u = -\delta(log\phi)_x \tag{7.155}$$

The initial conditions for the Burger's equation transform in a simple way into the initial conditions of the heat equation. Let us use the above transformation and see how it reduces to a heat equation. To start with, let us invoke the transformation

$$u = \psi_x \tag{7.156}$$

into the equation and integrate with respect to x, ignoring the function of integration. We have

$$\psi_t + (1/2)\psi_x^2 = \frac{\delta}{2}\psi_{xx} \tag{7.157}$$

Again the transformation

$$\psi = -\delta(\ln \phi) \tag{7.158}$$

reduces the above equation to

$$\phi_t = \frac{\delta}{2}\phi_{xx} \tag{7.159}$$

which is the famous heat equation. At this juncture, it would be worth mentioning that Whitham[174] introduced a similar transformation for the KdV equation. His transformation was

$$\sigma u = 12(\ln \phi)_{xx} \tag{7.160}$$

Unlike Burger's equation which becomes a linear heat equation, the transformed KdV equation continues to be nonlinear and is given by

$$\phi(\phi_t + \phi_{xxx})_x - \phi_x(\phi_t + \phi_{xxx}) + 3(\phi_{xx}^2 - \phi_x\phi_{xxx}) = 0 \tag{7.161}$$

In a similar way, another nonlinear equation, called the Riccati equation,

© 2015 by Taylor & Francis Group, LLC

which is governed by the nonlinear equation, with a suitable transformation is transformed into a second-order linear equation. The Riccati equation is given by

$$u' = f(x) + g(x)u + h(x)u^2 \tag{7.162}$$

The above equation is an ordinary differential equation, unlike the Burger's and KdV equations. The transformation given by

$$u = -\frac{1}{h(x)}\frac{d}{dx}(\ln \Phi) \tag{7.163}$$

reduces the Riccati equation to

$$\Phi'' + p(x)\Phi = 0 \tag{7.164}$$

where $p(x)$ is a specific combination of the functions f, g, and h. It turns out that Burger's[66] arrived at this transformation by seeking similarity solutions:

$$u = t^{-1/2}S(z), \qquad z = (2\delta t)^{-1/2}x$$

Some properties of Burger's equation include: (i) shifting of the origin $-$ $x - x_0 \to x$, $t - t_0 \to t$, $u \to u$, $\phi \to \phi$, where x_0 and t_0 are arbitrary independent constants, leave Burger's equation invariant and (ii) $x/\alpha \to x$, $t/\alpha^2 \to t$, $\alpha u \to u$, $\beta\phi \to \phi$, where α and β are independent scale factors, the change of scale leave the equation invariant. The same holds for Galilean transformations.

The special case in Burger's equation when $\delta = 2$ is interesting as this admits similarity solutions. Consider the change in the variable

$$u = \frac{1}{(2t+m)^{1/2}}f(\eta), \qquad \eta = \frac{x}{(2t+m)^{1/2}} \tag{7.165}$$

where f satisfies the nonlinear equation

$$f'' + f'(\eta - f) + f = 0 \tag{7.166}$$

Another change in variable is given by

$$u = \frac{1}{\eta} + (t+d)^{-1}f(\eta), \qquad \eta = \frac{t+d}{x+R} \tag{7.167}$$

Here f satisfies the nonlinear equation

$$\eta^2 f'' + 2\eta f' + ff' = 0 \tag{7.168}$$

The solution of the above equation in explicit form is given by

$$f = a_2 \tanh\Big[\frac{a_2}{2}(a_3 - \eta^{-1})\Big] \tag{7.169}$$

The m, d, R, a_2, and a_3 are all arbitrary constants.

© 2015 by Taylor & Francis Group, LLC

Consider the initial value problem for Burger's equation:

$$u(x,0) = f(x), \qquad -\infty < x < \infty \tag{7.170}$$

Using the Cole−Hopf transformation in an integrated form (ignoring the function of integration), we have

$$\phi(x,t) = \exp\Big[-\frac{1}{\delta}\int^{x} u(\xi,t)d\xi\Big] \tag{7.171}$$

so that the appropriate initial condition for the heat equation may be written as

$$\begin{aligned}\phi(x,0) &= \exp\Big[-\frac{1}{\delta}\int_0^{x} u(\xi,0)d\xi\Big] \\ &= \exp\Big[-\frac{1}{\delta}\int_0^{x} f(\xi)d\xi\Big] = \Phi(x)\end{aligned} \tag{7.172}$$

The existence and uniqueness of the solution may be ascertained subject to the following conditions: $f(x)$ is integrable in every finite x interval and

$$\int_0^{x} f(\xi)d\xi = 0(x^2) \tag{7.173}$$

and for $|x|$ large, the solution may be written as

$$u(x,t) = \frac{\int_{-\infty}^{\infty} \frac{x-\xi}{t}\exp\Big[-\frac{1}{\delta}F(x,\xi,t)\Big]d\xi}{\int_{-\infty}^{\infty}\exp\Big[-\frac{1}{\delta}F(x,\xi,t)\Big]d\xi} \tag{7.174}$$

where

$$F(x,\xi,t) = \frac{(x-\xi)^2}{2t} + \int_0^{\xi} f(\eta)d\eta \tag{7.175}$$

is a regular solution of Burger's equation in the half plane $t > 0$, satisfying the initial condition

$$\int_0^{x} u(\xi,t)d\xi \to \int_0^{a} f(\xi)d\xi, \qquad as \quad x \to a, \quad t \to 0 \tag{7.176}$$

for every a. If, in addition, $f(\xi)$ is continuous at $x = a$, then

$$u(x,t) \to f(a), \qquad as \quad x \to a, \quad t \to 0 \tag{7.177}$$

© 2015 by Taylor & Francis Group, LLC

7.11.2 Stationary Solutions and Shock Structure

Burger's equation may be classified as a quasi-linear equation, in the sense that it is linear in its highest derivative u_{xx}. Also, the coefficients of the derivative terms depend only on the dependent variable. Thus, one can look for traveling wave solutions depending on $x-Ut$ only, where U is the (constant) speed of the wave. Substituting $\xi = x - Ut$ into Burger's equation and looking for shock-like solutions with $u \to u_1, u_2$ as $\xi \to \pm\infty$, where $u_2 > u_1$, we have

$$-Uu_\xi + uu_\xi = \frac{\delta}{2}u_{\xi\xi} \tag{7.178}$$

The solutions of the above equation assume constant values of u_1, and u_2 at $\xi = \pm\infty$. The derivative of u vanishes at $\xi \to \pm\infty$, so that the integration of the above equation gives

$$\frac{1}{2}u^2 - Uu + C = \frac{\delta}{2}u_\xi \tag{7.179}$$

where C is a constant. Further, if we impose the conditions at $\xi = \pm\infty$, we have

$$\frac{1}{2}u_1^2 - Uu_1 = \frac{1}{2}u_2^2 - Uu_2 = -C \tag{7.180}$$

The constants may be evaluated as follows:

$$U = \frac{1}{2}(u_1 + u_2), \qquad C = \frac{1}{2}u_1u_2 \tag{7.181}$$

Substituting the above values into equation (7.179) yields

$$(u - u_1)(u_2 - u) = -\delta u_\xi \tag{7.182}$$

Upon integration, the above equation reduces to

$$\frac{\xi}{\delta} = \frac{1}{u_2 - u_1}\ln\frac{u_2 - u}{u - u_1} \tag{7.183}$$

In a more explicit form, the solution can be written as

$$u = u_1 + \frac{u_2 - u}{1 + \exp\dfrac{u_2 - u}{\delta}(x - Ut)}, \qquad U = \frac{u_1 + u_2}{2} \tag{7.184}$$

The above solution describes the structure of a uniformly propagating shock with end conditions u_1 and u_2. The velocity of the shock U is the mean of the end velocities and is independent of the shock structure. The solution is also an intermediate asymptotic for a class of solutions arising out of different initial conditions with asymptotically same end conditions which converge as $t \to \infty$. An interesting consequence of the above equation is due to Lighthill[69] and

© 2015 by Taylor & Francis Group, LLC

deals with piecewise shockwave formation. For this, let us use the following conditions:

$$u(x,0) = f(x) = u_1, \quad x > 0, \qquad u_2, \quad x < 0 \tag{7.185}$$

where $u_2 > u_1$. The initial conditions in terms of ϕ reduce to

$$\phi(x,0) = e^{-u_1 x/\delta}, \quad x > 0, \qquad e^{-u_2 x/\delta}, \quad x < 0 \tag{7.186}$$

The solution in this case can be written as

$$u(x,t) = u_1 + \frac{u_2 - u_1}{1 + \left(\exp\left[\frac{u_2 - u_1}{\delta}\left(x - \frac{u_1 + u_2}{2}t\right)\right]\right)\frac{\int_{-(x-u_1 t)}^{\infty} e^{-y^2/2M} dy}{\int_{(x-u_2 t)}^{\infty} e^{-y^2/2M} dy}} \tag{7.187}$$

In the limit $t \to \infty$, such that $u_1 < x/t < u_2$, the lower limits of the integrals in the denominator of the above solution are both large and negative, so that the ratio of the integrals tends to 1. This solution tends to the traveling wave solution (7.184). This is referred to in the literature as the Taylor shock structure[70].

7.11.3 Single-Hump Solution

The traveling wave solution discussed above is often called the similarity solution of the first kind. There is another kind of similarity solution of the second type, which is the single-hump solution. In this case, one of the important parameters is the Reynolds number R, given by

$$R = \frac{1}{\delta}\int_{-\infty}^{\infty} u dx = \frac{A}{\delta} \tag{7.188}$$

Consider the similarity form

$$u = \left[\frac{\delta}{t}\right]^{1/2} f\left[\frac{x}{(\delta t)^{1/2}}; \frac{A}{\delta}\right] \tag{7.189}$$

This corresponds to the similarity solution of the heat equation

$$\phi = c_1 + c_2 \int_{x/(\delta t)^{1/2}}^{\infty} e^{-y^2/2} dy = c_1 + c_2 erfc \frac{x}{(2\delta t)^{1/2}} \tag{7.190}$$

which depends on the similarity variable $x/(\delta t)^{1/2}$ only. $erfc$ is the error function. The constants c_1 and c_2 are chosen in such a way that $u \to 0$ as $x \to \pm\infty$. We introduce the transformation

$$\phi(x,t) = \exp\left(\frac{1}{\delta}\int_x^{\infty} u dx\right) \tag{7.191}$$

© 2015 by Taylor & Francis Group, LLC

The constants c_1 and c_2 are given by

$$c_1 = 1, \qquad c_2 = \frac{e^R - 1}{\sqrt{(2\pi)}} \tag{7.192}$$

Using the Cole−Hopf transformation, the solution may be written as

$$\begin{aligned} u &= \left(\frac{\delta}{t}\right)^{1/2} \frac{e^{-x^2/2\delta t}}{\frac{\sqrt{2\pi}}{e^R - 1} + \int_{x/(\delta t)^{1/2}}^{\infty} e^{-y^2/2} dy} \\ &= \left(\frac{\delta}{t}\right)^{1/2} \frac{e^{-x^2/2\delta t}}{\frac{\sqrt{2\pi}}{e^R - 1} + \sqrt{(\pi/2)} erfc \frac{x}{(2\delta t)^{1/2}}} \end{aligned} \tag{7.193}$$

For large values of R, the solution has the asymptotic form

$$u \approx \frac{x}{t}$$

To summarize the single-hump solution, we introduce the following transformation:

$$u = (\frac{\delta}{t})^{1/2} [H(\xi)]^{-1} \tag{7.194}$$

This transforms Burger's equation into the nonlinear ordinary differential equation in the similarity variable $\xi = x/(2\delta t)^{1/2}$ as

$$HH'' - 2H'^2 + 2\xi HH' - 2\sqrt{2}H' - 2H^2 = 0 \tag{7.195}$$

Here the prime denotes differentiation with respect to ξ. This equation has an exact solution given by

$$H = \frac{\sqrt{(2\pi)}}{e^R - 1} e^{\xi^2} + \sqrt{(\pi/2)} e^{\xi^2} erfc(\xi) \tag{7.196}$$

7.11.4 Planar N-Wave

A solution which is more physically important as compared to the single-hump solution is the N wave. These waves which arise in nature, say, from a spherical explosion or an aircraft with idealized spherical or cylindrical symmetry, carry an important geometrical effects due to the shape of the source. The N waves, with discontinuous shocks at $|x| = 1$ on the front and the tail, have the following form:

$$u = u(x, t_1) = x, \quad for \ \ |x| < 1; \quad = 0, \quad for \ \ |x| > 1 \tag{7.197}$$

We look for a solution of the heat equation, which is even about the node of

© 2015 by Taylor & Francis Group, LLC

the wave, $x = 0$, so that the corresponding solution of Burger's equation, using the Cole−Hopf transformation is odd about $x = 0$ and vanishes at $x = \pm\infty$. Such a solution is given by

$$\phi = 1 + (\frac{t_0}{t})^{1/2} e^{-x^2/2\delta t} \tag{7.198}$$

where t_0 is a constant. The corresponding solution of Burger's equation is

$$u = -\delta \frac{\phi_x}{\phi} = \frac{x/t}{1 + (\frac{t}{t_0})^{1/2} e^{x^2/2\delta t}} \tag{7.199}$$

As mentioned earlier, we define the Reynolds number R as given below:

$$R = \frac{A}{\delta} = \frac{1}{\delta} \int_0^\infty u dx = \ln \phi(0, t) = \ln\Big[1 + (\frac{t_0}{t})^{1/2}\Big] \tag{7.200}$$

Unlike the single-hump solution, here R is not a constant and decays to zero. The above equation may be written as

$$u = \frac{x/t}{1 + e^{x^2/2\delta t}/(e^R - 1)} \tag{7.201}$$

For the initial stages of the propagation of the N wave, when $R \gg 1$, we may write the above equation as

$$u \approx \frac{x}{t}[1 + e^{x^2/2\delta t - R)}]^{-1} \tag{7.202}$$

so that

$$\begin{aligned} u &\approx \frac{x}{t}, \qquad |x| < (2\delta R t)^{1/2} \\ u &\approx 0, \qquad |x| > (2\delta R t)^{1/2} \end{aligned} \tag{7.203}$$

This is the nonviscous solution of Burger's equation. The shock center x_s itself may be found by locating the point such that u is midway between the maximum of the discontinuous profile and zero, i.e., when

$$1 + \frac{e^{x^2/2\delta t}}{(e^R - 1)} \approx 2$$

Thus, we have the expression for x_s as

$$x_s = (\delta \ln(t_0))^{1/2} t^{1/2} - \frac{1}{2} \left(\frac{\delta}{\ln(t_0)}\right)^{1/2} t^{1/2} \ln t \tag{7.204}$$

The first term on the right-hand side of the above equation is the shock law according to inviscid (weak) shock theory, while the second term is the effect of diffusion and is often referred to as the shock wave displacement due to dissipation.

© 2015 by Taylor & Francis Group, LLC

7.11.5 Backlund Transformation

Consider a transformation as given below:

$$p' = f(t, x, u, u', p, q), \qquad q' = \psi(t, x, u, u', p, q) \tag{7.205}$$

with $p' = \partial u'/\partial x$ and $q' = \partial u'/\partial t$ as derivatives of the new dependent variable u'. If we find the functions f and ψ such that the pair of equations given above (equation 7.205) is integrable if and only if u and u' satisfy the differential equations

$$r + q + H(p, u, x, t) = 0 \tag{7.206}$$

$$r' + q' + G(p', u', x, t) = 0 \tag{7.207}$$

then the equations (7.205), define a Backlund transformation between the two equations (7.206 and 7.207),

Nimmo and Crighton[71] have shown that nonlinear equations which admit Backlund transformation are basically modifications of Burger's equation, namely, the inhomogeneous Burger's equation given by

$$u_t + uu_x = \delta u_{xx} + f(x, t) \tag{7.208}$$

Backlund transformations exist for KdV equations also. In passing, we mention that the nonlinear diffusion equation given by

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\Big[\frac{1}{(au+b)^2}\frac{\partial u}{\partial x}\Big] + \frac{d}{(au+b)^2}\frac{\partial u}{\partial x} \tag{7.209}$$

may be linearized by a combination of Cole−Hopf and Backlund transformations.

In this chapter, we have dealt with discontinuities in surfaces, the definition of Rankine−Hugoniot conditions, and the notion of planar, perpendicular, and oblique shocks. We have dealt with the concepts of weak shocks and waves in a polytropic gas. Shocks in collisionless plasmas and MHD were discussed, with an application to the Sun and Earth's bow shocks. Finally a section on nonlinear studies was included at the end of the chapter. In the next two chapters, we will have more to say on waves in optics and plasmas, before going to the final chapter on different types of instabilities.

7.12 Exercises

1. Using Rankine−Hugunoit conditions, show that

$$V_2^2 - V_1^2 = (p_1 - p_2)\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right)$$

Express the ratio of pressures in terms of Mach numbers.

© 2015 by Taylor & Francis Group, LLC

2. Discuss briefly the role of the self-similar variable $\xi = R(\frac{\rho_1}{Et^2})^{1/5}$ for the Taylor−Sedov blast wave solution.

3. Under what conditions, the solution of the equation $\rho_t + c(\rho)\rho_x = 0$ may be compared to the solution of the equation $\rho_t + c_0\rho_x = 0$.

4. Show that $u(x,t) = f(x-ut)$ is a solution of the equation $u_t + uu_x = 0$ with the condition $u(x,0) = f(x)$. Discuss the behaviour of $f(x)$.

5. Define magneto sonic Mach number as $M_{MS} = \frac{U_{SW}}{U_{MS}}$, where U_{SW}, and U_{MS} are the speed of solar wind and magneto sonic waves, respectively. For perpendicular shocks, calculate M_{MS}.

6. Using Rankine−Hugunoit relations and defining $\eta = \rho_2/\rho_1$, $\xi = p_2/p_1$ and $M = U_1/C_1$, show that the conservation of momentum relation across the shock can be written as

$$M_1^2\left(1 - \frac{1}{\eta}\right) = \frac{(\xi - 1)}{\gamma}$$

7. Consider the ratio of pressure p_2/p_1 given by

$$\frac{p_2}{p_1} = \frac{(\gamma+1)U_1 - (\gamma-1)U_2}{(\gamma+1)U_2 - (\gamma-1)U_1}$$

For $\gamma > 1$, derive the relation between U_1 and U_2.

8. The ideal gas equation is $p = \rho RT$, so that $T_2/T_1 = p_2\rho_1/p_1\rho_2$. Show that

$$\frac{p_2}{p_1} \to \frac{2\gamma(\gamma-1)}{(\gamma+1^2}M_1^2U_1^2$$

for large M_1^2.

9. Consider the equation

$$u_t + uu_x = \gamma u_{xx}$$

Find the solution of the above equation for $u \to c_1$ as $x \to -\infty$ and $u \to c_2$ as $x \to +\infty$.

10. Consider the Riccati equation

$$u' = f(x) + g(x)u + h(x)u^2$$

Show that the transformation

$$u = -\frac{1}{h(x)}\frac{d}{dx}(\ln \Phi)$$

reduces the above equation to a linear one.

© 2015 by Taylor & Francis Group, LLC

Chapter 8

Waves in Optics

8.1 Optical Phenomena 354
8.1.1 Classical Optics 354
8.1.1.1 Geometrical Optics 354
8.1.1.2 Wave Optics 355
8.1.2 Modern Optics 355
8.1.2.1 Quantum Optics 355
8.1.2.2 Statistical Optics 357
8.1.3 Velocity of Light 358
8.2 Nonmonochromatic Fields 359
8.2.1 Complex Representation 360
8.2.2 Power Spectrum 363
8.2.3 Notion of Convolution 363
8.2.3.1 Relationship 364
8.2.3.2 Orthogonality 366
8.2.4 Quasi-Monochromatic Fields 366
8.3 Emission of Wave-Trains 368
8.3.1 Coherence Length and Coherence Time 372
8.3.2 Damped Harmonic Vibrations 375
8.4 Polarization of Plane Monochromatic Waves 377
8.4.1 Instantaneous Optical Field and Polarization Ellipse 379
8.4.2 Stokes Parameters 383
8.4.2.1 Poincaré Sphere 386
8.4.2.2 Degree of Polarization 387
8.4.3 Measurements of Stokes Parameters 389
8.4.3.1 Polarizer 390
8.4.3.2 Retarder 392
8.4.3.3 Rotator 393
8.4.4 Stokes Intensity Formula 394
8.4.5 Jones Matrix 395
8.4.5.1 Jones Matrix for a Polarizer 396
8.4.5.2 Jones Matrix for a Retarder 399
8.4.5.3 Jones Matrix for a Rotator 400
8.4.6 Mueller Matrix 401
8.4.6.1 Mueller Matrix of a Polarizer 402
8.4.6.2 Mueller Matrix of a Retarder 403
8.4.6.3 Mueller Matrix of a Rotator 404
8.4.7 Rotated Polarizing Elements 405
8.4.7.1 Quarter-Wave Retarder 406
8.4.7.2 Half-Wave Retarder 407
8.4.8 Elliptical and Circular Polarizer and Analyzer 407
8.4.8.1 Left-Handed Circular Polarizer 408
8.4.8.2 Left-Handed Circular Analyzer 408
8.4.8.3 Left Circular Analyzer and Polarizer 409
8.4.9 Polarimetry 409
8.4.9.1 Polarimeter 409
8.4.9.2 Astronomical Polarimeter 410
8.5 Exercises 412

© 2015 by Taylor & Francis Group, LLC

8.1 Optical Phenomena

Optical science concerns the genesis and propagation of light, the changes it undergoes, and other phenomena associated with the electromagnetic mode. According to the excerpts from statutes of the International Commission for Optics (ICO), optics is the field of science and engineering encompassing the physical phenomena and technologies associated with the generation, transmission, manipulation, detection, and utilization of light. It extends on both sides of the visible part of the electromagnetic spectrum as far as the same concepts apply.

Observations of the behavior of light have revealed that light travels as waves. The phenomena of these waves and their oscillatory motion possess many degrees of freedom. All electromagnetic phenomena can be assumed to follow from the Maxwell equations, which are based on the experimental validation. The union of electromagnetic theory with optics began with these equations describing waves traveling at the velocity of light. Hertz[25] demonstrated the existence of electromagnetic waves by developing an apparatus to produce and detect ultra/very high frequency radio waves. There are two major branches of optics, which are enumerated below in brief.

8.1.1 Classical Optics

Geometrical and physical optics are referred to as the classical optics. The classical model of light describes it as a transverse, electromagnetic wave. At the end of the nineteenth century, the wave nature of light was confirmed when it was applied to explain the phenomena of reflection, refraction, interference, diffraction, and polarization.

8.1.1.1 Geometrical Optics

Geometrical optics deals with the image formation and related phenomena that can be discussed within the framework of the laws concerning reflection, refraction, and rectilinear propagation. It is the study of optics where light propagation is described using ray tracing. It applies in those cases where interference and diffraction are ignored. It has found an application in the method of ray tracing widely used in computer graphics.

Geometrical optics commenced with the discovery of the law of refraction by Willebrord Snell (1580−1626) in 1621 and its description in mathematical terms by René Descartes (1596−1650) in 1637. In 1662, Pierre de Fermat (1601−1665) showed that the law of refraction could be obtained from the principle of least action, also known as the Fermat law of least time, which manifests itself by way of the fact that the actual ray path between two points takes the least time. Later, during 1665−1666, Issac Newton (1642−1727) conducted the refraction experiments, which led to his theory of

© 2015 by Taylor & Francis Group, LLC

colors. Polarization of light by reflection was discovered in 1808 by Étienne-Louis Malus (1775−1812) and the polarizing angle was discovered by David Brewster (1781−1868) in 1811.

8.1.1.2 Wave Optics

Wave optics (or physical optics) deals primarily with the nature and properties of light itself. Founded on the Maxwell equations, it describes diffraction and interference phenomena at length. As stated in Section 1.6.2, the concept of the interference of waves, developed in mechanics in the eighteenth century, was introduced into optics by Young[34]. He observed that the superposition of waves (see Section 1.4.2) was insufficient to describe the phenomenon of optical interference which could not lead to the observed interference pattern. In the eighteenth century, Euler, d'Alembert, and Lagrange developed the wave equation from the Newtonian mechanics and studied its consequences, for instance, propagating and standing waves (see Section 1.4.3). In order to describe the interference pattern, Young used the concept of energy from mechanics. The mechanical developments of that period were crucial to the development of optics in the early nineteenth century. The wave equation (see Section 1.3.1) that arose in mechanics was subsequently applied to optics ([167], [105]). Following the description of the effects of diffraction in 17th century by Huygens, and the introduction of the principle of interference by Young, wave optics was developed.

8.1.2 Modern Optics

Modern optics encompasses the areas of optical science which are elucidated below.

8.1.2.1 Quantum Optics

The advancement of the theory of light progressed rapidly after the initiation of quantum theory. Quantum mechanics deals with the unified description. Quantum optics uses the fact that the electromagnetic field can interact with its surroundings in discrete energy, $h\nu$, or multiples thereof, and that energy differences in atomic jumps are emitted in these discrete amounts of energy (quantum atomic levels) as electromagnetic radiation[141]. A laser is the main tool for the coherent optics.

The quantum optics deals with absorption and emission, which cannot be described with just the Maxwell equation. This subject began in 1818 with the studies of the absorption lines of the elements by Fraunhofer leading through the Planck energy quantization[29], which appeared to be incompatible with the wave theory of light. While interpreting the spectrum of electromagnetic radiation emitted by a black body, Planck postulated that the oscillating electric charges which gives rise to light emission can have discrete energies. His hypotheses are

© 2015 by Taylor & Francis Group, LLC

- an oscillator can have one of a discrete set of possible values of energy, $nh\nu$, in which n is a whole number, ν the frequency of the radiation, and the Planck constant $h(= 6.626196 \times 10^{34} Joules\,\text{J}.s = 4.136 \times 10^{-15} eVs)$, and
- the emission and absorption of radiation are associated with transitions between these levels.

The radiation of frequency, ν, can exchange energy with matter in units (quanta) of $h\nu$, which is the energy separation between the levels. When an oscillator emits a photon, it drops from energy, $nh\nu$, to a level $(n-1)h\nu$.

1. Uncertainty principle: The Planck constant also occurs in statements of the Werner Heisenberg (1901−1976) uncertainty principle of quantum mechanics, which states that certain sets of conjugate observables cannot be known simultaneously with infinite precision. It provides a lower bound on the product of the standard deviations of two conjugate variables, such as position and momentum for a system, the angular momentum and angular displacement for a system, or the energy and time for a moving particle[72]. It is customary to choose the conjugate variables as position and momentum since their product has the dimension of action, i.e., energy multiplied by time. A pair of physical variables describing a quantum-mechanical system, such that their commutator is a nonzero constant, are also known as complementary variables. The precision in measurement of one of them destroys the possibility of measurement of the other. Uncertainty should not be construed as a limitation of measurement on the measuring device; it is an inadequacy inbuilt in the process of measurement. It is improbable to pinpoint the exact position and exact momentum of an object at the same time. For example, a moving body corresponds to a single wave group, which can be thought of in terms of the superposition of trains of harmonic waves. However, an infinite number of wave trains with different frequencies, wave numbers, and amplitudes is needed for an isolated group of arbitrary shape. The narrower the wave group, the greater the range of wavelengths involved, which means a well-defined position (Δx), but a large uncertainty in the momentum (Δp) of the particle. On the contrary, a wide group means a more precise momentum with a less precise position.

2. Photon: Following the suggestion, made by Planck, Einstein[73] put forward that light consists of a finite number of energy quanta, called *lichtquanten*, which are localized at points in space. According to him, the observations associated with black body radiation could be understood if the energy of light were discontinuously distributed in space. This gave way to a term *photon* introduced by a chemist Gilbert in 1925, which eventually led to the development of quantum theory. A photon is a quantum or discrete bundle of electromagnetic energy, which occurs when an electron jumps from a higher energy state to lower one.

© 2015 by Taylor & Francis Group, LLC

This process is known as spontaneous emission of radiation. Its energy depends on the radiation frequency of the light; the shorter the wavelength, the more energy each of these quanta has, the longer the wavelength the smaller the energy of the quanta. There are photons of all energies from high-energy gamma rays to low-energy infrared and radio waves. A beam of light can be envisaged as a stream of photons that travel at the speed of light. They move without dividing and can only be produced and absorbed as complete units.

3. Wave-particle duality: Wave and particle natures are dual. When one is exhibited, the other remains suppressed. If radiation can exhibit particle nature, a particle can also show wave nature. From the wave nature of particle, de Broglie showed that the wavelength of a particle is directly related to the Planck constant and inversely proportional to the mass and velocity of the particle. The quantum theory of atomic structure, proposed by Niels Henrik David Bohr (1885−1962) in 1913[74], followed by the discovery of dual nature of electromagnetic radiation (wave-particle duality), predicted by de Broglie, led to the development of quantum mechanics. Erwin Schrödinger (1887−1961) developed the wave-particle duality equation[75].

8.1.2.2 Statistical Optics

With the statistical interpretation of quantum mechanics introduced by Max Born (1882−1970), statistical optics became a new branch of optics. Light is quantized, and the statistics of the photons results in uncertainties of the amplitude and the phase of the light field. A quantum state can be occupied by several particles, what is referred to as quantum statistics. This applies to indistinguishable particles. Born formulated the now-standard interpretation of the probability density function for $\Psi\Psi^*$ in the Schrödinger equation of quantum mechanics. To note, the probability distribution (or density) assigns to every interval of real numbers a probability, so that the probability axioms are satisfied. In quantum mechanics, a probability amplitude is, in general, a complex number whose modulus squared represents a probability or probability density. For example, the values taken by a normalized wave function Ψ are amplitudes, since $\Psi\Psi^* = |\Psi(x)|^2$ provides the probability density at position x. Probability amplitudes, defined as complex-number-valued function of position, may also correspond to probabilities of discrete outcomes. The probability of finding the particle described by the wave function (e.g., an electron in an atom) at that point is proportional to the square of the absolute value of the probability amplitude. That the physical meaning of the wave function is probabilistic was proposed by Max Born.

Statistical optics is the study of the properties of random light. The randomness of photons in fluctuating light fields, which may have different frequencies, arises because of unpredictable fluctuations of the light source or of

© 2015 by Taylor & Francis Group, LLC

the medium through which the light propagates. The randomness in light may also be generated by scattering from a rough surface or by turbulent fluids imparting random variations to the optical wavefront. Wolf[76] introduced a broad framework for considering the coherence properties of waves. Statistical optics gained impetus after the invention of laser in 1960.

8.1.3 Velocity of Light

A number of observations had been made towards the end of the nineteenth century establishing the velocity of light independent of the velocity of the source or of the observer. This result could not be explained by classical theory as it assumes that the physical quantities have a continuous range of values. In an article by Einstein[77], which dealt with the motion of small particles suspended in liquids, as required by molecular kinetic theory of heat by explaining Brownian motion[78], he showed that by measuring the mean squared displacement, the Avogadro number could be determined. This provided a definite proof for the existence of atoms. He extended and improved the Maxwell theory by modifying the foundations of classical mechanics to remove the apparent contradictions between mechanics and electrodynamics, which led to the introduction of a special theory of relativity. This theory has completely changed the concepts of space and time. He showed that space and time are closely linked and affect each other. Space, time, simultaneity, and mass are all relative, but the speed of light is absolute being independent of the notion of the source and the observer. Another major consequence of this theory was to establish the equivalence of mass and energy. Subsequently, Einstein[79] explained the varying kinetic energies of an object emitting radiation by considering the amount of change in kinetic energy of objects, as being equivalent to the change in its mass, that is,

$$E = \gamma m_0 c^2 = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}} \tag{8.1}$$

with $\gamma(\equiv \frac{1}{\sqrt{1-v^2/c^2}})$ as the Lorentz factor, m_0 the particle's rest mass, v the velocity of the particle, and c the speed of the light in vacuum; the greater the particle's momentum, the shorter its wavelength.

The relativistic mass, m, which is a function of the rest mass and the velocity of the object, of a particle is given by

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} \tag{8.2}$$

therefore, the relativistic momentum, $\mathbf{p}$, is written as

$$\mathbf{p} = \frac{m_0 v}{\sqrt{1 - v^2/c^2}} = \gamma m_0 \mathbf{v} \tag{8.3}$$

The total energy and momentum are conserved in an isolated system, and

© 2015 by Taylor & Francis Group, LLC

the rest energy of a particle is invariant. On applying the special theory of relativity, developed by Einstein, the relation between energy and momentum (p) of a particle is derived with equations (8.1 and 8.3):

$$E = \sqrt{(\mathbf{p}c)^2 + (m_0 c)^2} \tag{8.4}$$

in which $\mathbf{p}(= \hbar\boldsymbol{\kappa} = m\mathbf{c})$ is the momentum, defined by the product of the mass of a particle and its linear velocity, $\boldsymbol{\kappa}$ the wave vector, and $\hbar = h/(2\pi)$.

Since $m_0 c^2$ is invariant, so is $E^2 - p^2 c^2$; this quantity for a particle has the same value in all frame of reference. Later, Paul A. M. Dirac (1902−1984) developed a form of the Schrödinger wave equation that is consistent with the special theory of relativity[80] describing the behavior of fermions and predicted the existence of antimatter, which was identified by Anderson[81]. The Dirac quantization of an electromagnetic field as a collection of harmonic oscillators amid particles that can be created and annihilated was the first step toward quantum electrodynamics and, more generally, toward quantum field theories.

8.2 Nonmonochromatic Fields

In practice, the optical fields are never completely monochromatic. The electromagnetic waves emitted by the atoms are discharged as wave trains. Owing to the finite length of these wave-trains, the radiation forms a frequency spectrum. For a monochromatic wave field, the amplitude of vibration at any point is constant and the phase varies linearly with time. Conversely, the amplitude and phase in the case of a quasi-monochromatic wave field undergo irregular fluctuations[105]. The fluctuations arise since the real valued wave field $U^{(r)}$ consists of a large number of contributions that are independent of each other, the superposition of which gives rise to a fluctuating field. The rapidity of fluctuations depends on the light crossing time of the emitting region. The realistic light beam, $U^{(r)}$, that is regarded as a member of an ensemble consisting of all realizations of the field, fluctuates as a function of time. At optical frequencies, the fluctuating field components are not observable quantities, but are quadratic averages of them. Since, as a rule, the stochastic field is treated as ergodic, the ensemble average can be replaced by a time average. Let us consider that the waves (see equation 1.16), propagate in space where no sources and boundaries are present. There is a continuous multitude of such plane waves indexed by the continuous vector $\boldsymbol{\kappa}$. We have seen in Section 1.4.2 that any superposition of solutions of an equation is also a solution of the wave equation.

© 2015 by Taylor & Francis Group, LLC

8.2.1 Complex Representation

A nonmonochromatic wave can be expanded as a sum of monochromatic waves by using conventional Fourier analysis. An arbitrary function of time, such as the wave function, $U(\mathbf{r},t)$, at a fixed position $\mathbf{r}$, may be represented in the form of superposition of a large number of sinusoidal components of monochromatic waves of different frequencies, each component having a steady amplitude and phase over the period of observation, but both the amplitude and phase being random with respect to each other component. Hence, we may express

$$U(\mathbf{r},t) = \int_{-\infty}^{\infty} \widehat{U}(\mathbf{r},\nu) e^{i2\pi\nu t} d\nu \tag{8.5}$$

By carrying out the Fourier transform (see Section 1.3.3) of equation (8.5), we get

$$\widehat{U}(\mathbf{r},\nu) = \int_{-\infty}^{\infty} U(\mathbf{r},t) e^{-i2\pi\nu t} dt \tag{8.6}$$

By splitting the integral on the right-hand side (RHS) of equation (8.5) into two integrals with limits of $-\infty \to 0$ and $0 \to \infty$, it is simplified as

$$\int_{-\infty}^{0} \widehat{U}(\mathbf{r},\nu) e^{i2\pi\nu t} d\nu = \int_{0}^{-\infty} \widehat{U}^*(\mathbf{r},\nu) e^{-i2\pi\nu t} d\nu$$

in which $*$ stands for complex conjugate, so that $U(\mathbf{r},t)$ is the sum of a complex function and its conjugate[1]

$$U(\mathbf{r},t) = \int_{0}^{\infty} \left[\widehat{U}(\mathbf{r},\nu) e^{i2\pi\nu t} + \widehat{U}^*(\mathbf{r},\nu) e^{-i2\pi\nu t} \right] d\nu \tag{8.7}$$

The expression for the complex wavefunction, also called the analytic signal associated with the real function, $U^{(r)}(\mathbf{r},t)$, in the case of polychromatic wave is derived by analogy to the monochromatic case (see Section 1.4) yielding

$$U(\mathbf{r},t) = 2\int_{0}^{\infty} \widehat{U}(\mathbf{r},\nu) e^{i2\pi\nu t} d\nu \tag{8.8}$$

The complex wavefunction can be obtained from the wavefunction by (i) determining its FT, (ii) eliminating negative frequencies and multiplying by 2, and (iii) determining the inverse FT. Since each of its Fourier components satisfies the wave equation, the complex wavefunction $U(\mathbf{r},t)$ satisfies the wave equation.

The disturbance produced by a real physical source is calculated by the

[1] Let us consider that Z is a complex number:

$$Z + Z^* = 2\Re[Z]$$

hence, $\widehat{U}(-\nu)e^{-i2\pi\nu t}$ and $\widehat{U}(\nu)e^{i2\pi\nu t}$ are conjugate; therefore, $\widehat{U}(-\nu) = \widehat{U}^*(\nu)$.

© 2015 by Taylor & Francis Group, LLC

integration of the monochromatic signals over an optical band pass. For a real, nonmonochromatic vibration, $U^{(r)}(\mathbf{r},t)(-\infty \leq t \leq \infty)$ is expressed as

$$U^{(r)}(\mathbf{r},t) = \int_0^\infty a(\mathbf{r},\nu)\cos\left[2\pi\nu t - \psi(\mathbf{r},\nu)\right] d\nu \tag{8.9}$$

where $a(\mathbf{r},\nu)$ and $\psi(\mathbf{r},\nu)$ modulo (2π) are real functions, which provide the amplitude and phase, respectively, of each monochromatic component of frequency ν.

Equation (8.9) is the Fourier cosine integral representation of the real valued signal $U^{(r)}(\mathbf{r},t)$. The associated Fourier sine integral function is given by

$$U^{(i)}(\mathbf{r},t) = \int_0^\infty a(\mathbf{r},\nu)\sin\left[2\pi\nu t - \psi(\mathbf{r},\nu)\right] d\nu \tag{8.10}$$

Invoking the Euler formula, we derive the complex analytic signal, $U(\mathbf{r},t)$, associated with the real function, $U^{(r)}(\mathbf{r},t)$, as

$$U(\mathbf{r},t) = U^{(r)}(\mathbf{r},t) + iU^{(i)}(\mathbf{r},t) \tag{8.11}$$

The imaginary part $U^{(i)}(\mathbf{r},t)$ contains no new information about the optical field. Therefore, the complex function, $U(\mathbf{r},t)$, is derived in the form of a Fourier integral as

$$U(\mathbf{r},t) = \int_0^\infty a(\mathbf{r},\nu)e^{i[2\pi\nu t - \psi(\mathbf{r},\nu)]}d\nu \tag{8.12}$$

The signal in equation (8.12) contains positive frequencies. The real and imaginary parts of the frequency response of any physical system are related to each other by Hilbert transform[150]; this relationship is also known as Kramers−Kronig relationship. Hilbert transform $\widehat{f}(t)$ of a signal $f(t)$ is defined as

$$\widehat{f}(t) = f(t) \star \frac{1}{\pi t} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(\tau)}{t-\tau}d\tau \tag{8.13}$$

where $\star$ denotes the convolution.

The integral (equation 8.13) has the form of a convolution integral. Figure 8.1 depicts the Hilbert transform of a rectangular and a sinusoidal vibration. The integrals of equations (8.9 and 8.10) are allied integrals. Hence, the conjugate functions $U^{(r)}$ and $U^{(i)}$ are expressed as

$$\begin{aligned} U^{(r)}(t) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{U^{(i)}(\tau)}{t-\tau}d\tau \\ U^{(i)}(t) &= -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{U^{(r)}(\tau)}{t-\tau}d\tau \end{aligned} \tag{8.14}$$

Let $\widehat{U}^{(r)}(\mathbf{r},\nu)$ be the Fourier transform of $U^{(r)}(\mathbf{r},t)$; therefore, the Fourier

© 2015 by Taylor & Francis Group, LLC

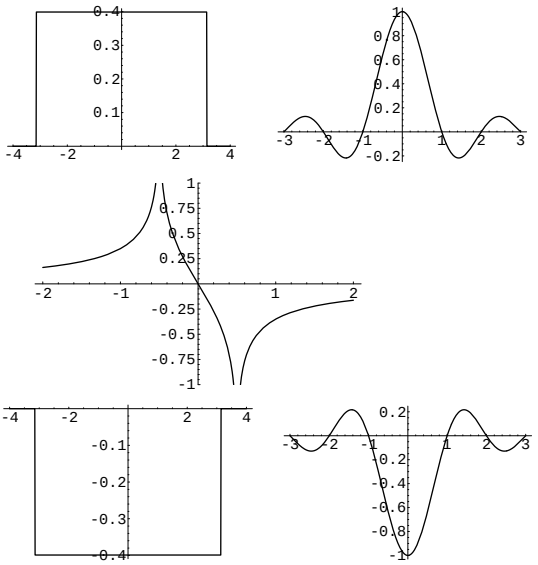


FIGURE 8.1: Hilbert transform of a rectangular as well as a sinusoidal vibration.

transform pair can be represented as

$$U^{(r)}(\mathbf{r},t) = \int_{-\infty}^{\infty} \widehat{U}^{(r)}(\mathbf{r},\nu)e^{i2\pi\nu t}d\nu \tag{8.15}$$

$$\widehat{U}^{(r)}(\mathbf{r},\nu) = \int_{-\infty}^{\infty} U^{(r)}(\mathbf{r},t)e^{-i2\pi\nu t}dt \tag{8.16}$$

Following equation (8.7), we may write equation (8.15) as

$$U^{(r)}(\mathbf{r},t) = 2\Re\left[\int_{0}^{\infty} \widehat{U}^{(r)}(\mathbf{r},\nu)e^{i2\pi\nu t}d\nu\right] \tag{8.17}$$

Comparing equations (8.9 and 8.17), we obtain

$$\widehat{U}(\mathbf{r},\nu) = \frac{1}{2}a(\mathbf{r},\nu)e^{-i\psi(\mathbf{r},\nu)} \qquad \nu \geq 0 \tag{8.18}$$

Therefore, equation (8.12) takes the same form as equation (8.8). Hence, $U(\mathbf{r},t)$ may be derived from $U^{(r)}(\mathbf{r},t)$ if the operations on the latter are linear. The real part provides the real-valued wave field according to equation (8.8) by representing $U^{(r)}$ as a Fourier integral. The following relationships follow from equations (8.8 and 8.15) by the Parseval theorem (see Section 8.2.2) and

© 2015 by Taylor & Francis Group, LLC

by the use of the relation (equation 8.11)[105]:

$$\begin{aligned}\int_{-\infty}^{\infty} \mid U^{(r)}(\mathbf{r},t) \mid^2 dt &= \int_{-\infty}^{\infty} \mid U^{(i)}(\mathbf{r},t) \mid^2 dt = \frac{1}{2}\int_{-\infty}^{\infty} U(\mathbf{r},t)U^*(\mathbf{r},t)dt \\ &= \int_{-\infty}^{\infty} \mid \widehat{U}(\mathbf{r},\nu) \mid^2 d\nu = 2\int_{0}^{\infty} \mid \widehat{U}(\mathbf{r},\nu) \mid^2 d\nu \quad (8.19)\end{aligned}$$

8.2.2 Power Spectrum

The power spectrum, or Parseval theorem is, in general, interpreted as a statement of conservation of energy; for instance, if $\mathcal{F}[f(t)] = \widehat{f}(\nu)$, we get

$$\begin{aligned}\int_{-\infty}^{\infty} \mid \widehat{f}(\nu) \mid^2 d\nu &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(t)e^{i2\pi\nu t}dt \int_{-\infty}^{\infty} f^*(t')e^{-i2\pi\nu t'}dt'\right]d\nu \\ &= \iint_{-\infty}^{\infty} f(t)f^*(t')\left[\int_{-\infty}^{\infty} e^{i2\pi\nu(t-t')}d\nu\right]dtdt' \\ &= \iint_{-\infty}^{\infty} f(t)f^*(t')\delta(t'-t)dtdt' = \int_{-\infty}^{\infty} |f(t)|^2 dt \quad (8.20)\end{aligned}$$

Equation (8.20) states that the integral of the squared modulus of a function is equal to the integral of the squared modulus of its spectrum. Also known as the Rayleigh theorem, it says that the total energy in the real domain is equal to the total energy in the Fourier domain. Figure 8.2 depicts the power spectrum of a sinusoidal function.

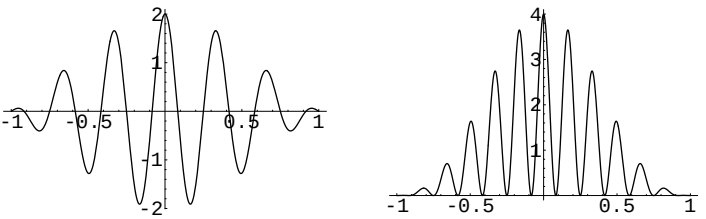


FIGURE 8.2: Left panel: a sinusoidal function, and right panel: its power spectrum.

8.2.3 Notion of Convolution

Convolution describes the action of an observing instrument when it takes a weighted mean of some physical quantity over a narrow range of some variable. When the form of the weighting function does not change appreciably as

© 2015 by Taylor & Francis Group, LLC

the central value of the variable changes, the observed quantity is the convolution of the distribution of the desired quantity with the weighting function, rather than the value of the desired quantity itself. The convolution of two functions simulates phenomena where each point is replaced by a spread function (see Figure 8.3), which is defined as

$$\begin{aligned} h(t) &= \int_{-\infty}^{\infty} f(t')g(t-t')dt' \\ &= f(t) \star g(t) \end{aligned} \tag{8.21}$$

in which $f(t)$ represents an input curve, $g(t)$ a blurring function, $h(t)$ the output value, and $\star$ the convolution parameter.

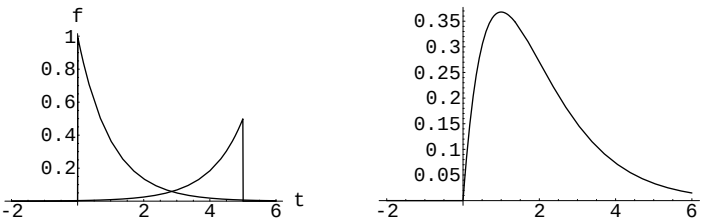


FIGURE 8.3: Convolution of two truncated exponentials.

The convolution of two functions in the space domain is equivalent to the more simple operation of multiplying their individual transforms

$$\widehat{h}(\nu) = \widehat{f}(\nu).\widehat{g}(\nu) \tag{8.22}$$

In the case of electrical filters, transfer functions are functions of frequency, while in the case of spectroscopy in which the scanning function is a function of frequency, the transfer function is a function of time. Let the Fourier integral representation of the real function $U^{(r)}$ be represented by

$$U^{(r)}(t) = \int_{-\infty}^{\infty} \widehat{U}^{(r)}(\nu)e^{-i2\pi\nu t}d\nu \tag{8.23}$$

and by the Fourier inversion formula

$$\widehat{U}^{(r)}(\nu) = \int_{-\infty}^{\infty} U^{(r)}(t)e^{i2\pi\nu t}dt \tag{8.24}$$

8.2.3.1 Relationship

As discussed in Section 8.2.1, the real field variable at a point is represented by a position vector $\mathbf{r}(= x, y, z)$, at time t. It is convenient to carry out the analysis in terms of associated analytic signal $U(\mathbf{r}, t)$ instead of the real field

© 2015 by Taylor & Francis Group, LLC

variable. The convolution relationship of two analytical signals is expressed as[122]

$$\begin{aligned}\int_{-\infty}^{\infty} U_1(t+\tau)U_2^*(t)dt &= \int_{-\infty}^{\infty} U_2^*(t)\left[2\int_0^{\infty}\widehat{U}_1(\nu)e^{i2\pi\nu(t+\tau)}d\nu\right]dt \\ &= 2\int_0^{\infty}\widehat{U}_1(\nu)e^{i2\pi\nu\tau}\left[\int_{-\infty}^{\infty}U_2^*(t)e^{i2\pi\nu t}dt\right]d\nu \end{aligned} \tag{8.25}$$

where $U_1(t)$ and $U_2(t)$ are the analytic signals associated with the real vibrations $U_1^{(r)}(t)$ and $U_2^{(r)}(t)$.

According to equation (8.8)

$$\begin{aligned} 2\widehat{U}_1(\nu) &= \int_{-\infty}^{\infty} U_2(t)e^{-i2\pi\nu t}dt \\ 2\widehat{U}_2^*(\nu) &= \int_{-\infty}^{\infty} U_2^*(t)e^{i2\pi\nu t}dt \end{aligned} \tag{8.26}$$

Therefore, equation (8.25) turns out to be

$$\int_{-\infty}^{\infty} U_1(t+\tau)U_2^*(t)dt = 4\int_0^{\infty}\widehat{U}_1(\nu)\widehat{U}_2^*(\nu)e^{i2\pi\nu\tau}d\nu \tag{8.27}$$

The convolution of the two functions $U_1^{(r)}(t)$ and $U_2^{(r)}(t)$, as well as the convolution of the functions $U_1^{(i)}(t)$ and $U_2^{(i)}(t)$ or $U_1^{(i)}(t)$ and $U_2^{(r)}(t)$, can be reduced as

$$\int_{-\infty}^{\infty} U_1^{(r)}(t+\tau)U_2^{(r)}(t)dt = \int_{-\infty}^{\infty} U_1^{(i)}(t+\tau)U_2^{(i)}(t)dt \tag{8.28}$$

$$\int_{-\infty}^{\infty} U_1^{(r)}(t+\tau)U_2^{(i)}(t)dt = -\int_{-\infty}^{\infty} U_1^{(i)}(t+\tau)U_2^{(r)}(t)dt \tag{8.29}$$

According to equation (8.11)

$$\begin{aligned}\int_{-\infty}^{\infty} U_1(t+\tau)U_2^*(t)dt &= \int_{-\infty}^{\infty}\left[U_1^{(r)}(t+\tau)+iU_1^{(i)}(t+\tau)\right] \\ &\quad\times\left[U_2^{(r)}(t)-iU_2^{(i)}(t)\right]dt \end{aligned} \tag{8.30}$$

By using equations (8.28 and 8.29), we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} U_1(t+\tau)U_2^*(t)dt &= 2\int_{-\infty}^{\infty} U_1^{(r)}(t+\tau)U_2^{(r)}(t)dt \\ &\quad -2i\int_{-\infty}^{\infty} U_1^{(r)}(t+\tau)U_2^{(i)}(t)dt \end{aligned} \tag{8.31}$$

© 2015 by Taylor & Francis Group, LLC

8.2.3.2 Orthogonality

Equation (8.31) states that the real part of the convolution of two analytic signals equals up to a factor of 2, the convolution of the real functions associated with the analytic signals. By setting $U_1(t) = U_2(t)$ in equation (8.31)

$$\begin{aligned}\int_{-\infty}^{\infty} U_1(t+\tau)U_1^*(t)dt &= 2\int_{-\infty}^{\infty} U_1^{(r)}(t+\tau)U_1^{(r)}(t)dt \\ &\quad -2i\int_{-\infty}^{\infty} U_1^{(r)}(t+\tau)U_1^{(i)}(t)dt \end{aligned} \tag{8.32}$$

From equation (8.29), we find

$$\int_{-\infty}^{\infty} U_1^{(r)}(t+\tau)U_1^{(i)}(t)dt = -\int_{-\infty}^{\infty} U_1^{(i)}(t+\tau)U_1^{(r)}(t)dt \tag{8.33}$$

and for $\tau = 0$:

$$\int_{-\infty}^{\infty} U_1^{(r)}(t)U_1^{(i)}(t)dt = 0 \tag{8.34}$$

It has been observed from equation (8.34) that the functions $U^{(r)}(t)$ and $U^{(i)}(t)$ are orthogonal. By setting $\tau = 0$ in equation (8.32), we find

$$\int_{-\infty}^{\infty} U_1(t)U_1^*(t)dt = 2\int_{-\infty}^{\infty} \mid U_1^{(r)}(t) \mid^2 dt \tag{8.35}$$

Hence, the time integral of the square of the modulus of the analytic signal is equal to twice the time integral of the real functions with which the analytic signal is associated. By adding $\tau = 0$ to equation (8.27), and using the Parseval theorem (see Section 8.2.2), we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} U_1(t)U_1^*(t)dt &= 2\int_{-\infty}^{\infty} \mid U^{(r)}(t) \mid^2 dt = 2\int_{-\infty}^{\infty} \mid U^{(i)}(t) \mid^2 dt \\ &= 2\int_{-\infty}^{\infty} \mid \widehat{U}(\nu) \mid^2 d\nu = 4\int_{0}^{\infty} \mid \widehat{U}(\nu) \mid^2 d\nu \end{aligned} \tag{8.36}$$

8.2.4 Quasi-Monochromatic Fields

A quasi-monochromatic wave has Fourier components with frequencies confined within a frequency interval of width $\Delta\nu$ surrounding a central frequency ν_0, such that $\Delta\nu \ll \nu_0$. Figure 8.4 depicts the Fourier transform of a sinusoidal vibration. Here, the interval $\Delta\nu$ of the frequencies between A and B is $\Delta\nu = \nu - \nu_0 = 1/\tau$. The analytic signal given by equation (8.8) can be expressed as

$$U(\mathbf{r},t) = 2e^{i2\pi\nu_0 t}\int_0^{\infty} \widehat{U}(\mathbf{r},\nu)e^{i2\pi(\nu-\nu_0)t}d\nu \tag{8.37}$$

© 2015 by Taylor & Francis Group, LLC

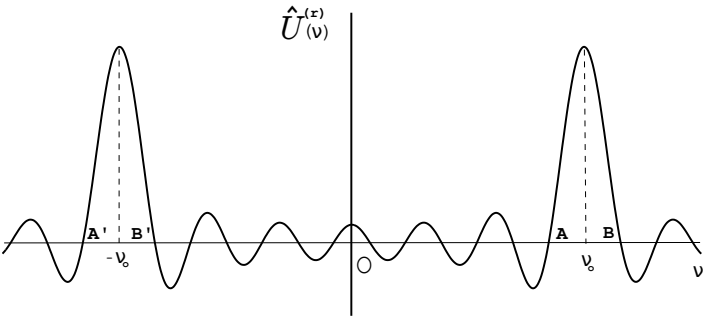


FIGURE 8.4: Fourier transform of a sinusoidal vibration.

Let us define a complex function, $a(\mathbf{r},t)[= A(\mathbf{r},t)e^{-i\psi(\mathbf{r},t)}]$, and therefore, the instantaneous complex amplitude of the vibration, $a(\mathbf{r},t)$, is written as

$$a(\mathbf{r},t) = 2\int_0^\infty \widehat{U}(\mathbf{r},\nu)e^{i2\pi(\nu-\nu_0)t}d\nu \tag{8.38}$$

therefore, equation (8.37) is recast as

$$U(\mathbf{r},t) = a(\mathbf{r},t)e^{i2\pi\nu_0 t} \tag{8.39}$$

If the Fourier transform $\widehat{U}(\mathbf{r},\nu)$ of $U^{(r)}(\mathbf{r},t)$ differs appreciably from zero nearly in the neighborhood of $\nu = \nu_0$, the integral (equation 8.38) represents a superposition of harmonic components of low frequencies. Moreover, if $\Delta\nu = \nu - \nu_0$ over which $\widehat{U}(\mathbf{r},\nu)$ differs from zero is small with respect to ν_0, the exponential term $e^{-i2\pi(\nu-\nu_0)t}$ varies slowly with respect to the term $e^{-i2\pi\nu_0 t}$ in the course of time. In this case, the light emitted is known to be a quasi-monochromatic source. In equation (8.39), the variations of $a(\mathbf{r},t)$ in time are slow with respect to the oscillatory term $e^{-i2\pi\nu_0 t}$. This equation is assumed to be a monochromatic vibration with a variable amplitude $a(\mathbf{r},t)$ and a frequency equal to ν_0. On replacing the value of $a(\mathbf{r},t)$ given by equation (8.38) in equation (8.39), we obtain

$$U(\mathbf{r},t) = A(\mathbf{r},t)e^{i[2\pi\nu_0 t-\psi(\mathbf{r},t)]} \tag{8.40}$$

Since $U^{(r)}$ and $U^{(i)}$ are the real and imaginary parts of U, in terms of A and ψ:

$$\begin{aligned} U^{(r)}(\mathbf{r},t) &= A(\mathbf{r},t)\cos[2\pi\nu_0 t - \psi(\mathbf{r},t)] \\ U^{(i)}(\mathbf{r},t) &= A(\mathbf{r},t)\sin[2\pi\nu_0 t - \psi(\mathbf{r},t)] \end{aligned} \tag{8.41}$$

These formulae (equation 8.41) express $U^{(r)}$ and $U^{(i)}$ in the form of normalized

© 2015 by Taylor & Francis Group, LLC

signals of carrier frequency ν_0. The complex analytic signal is intimately connected with the envelope of the real signal. By squaring these subequations, followed by the additions, we get

$$A^2(\mathbf{r},t) = \mid U^{(r)}(\mathbf{r},t) \mid^2 + \mid U^{(i)}(\mathbf{r},t) \mid^2 \tag{8.42}$$

Thus, in terms of the analytic signal U, the envelope $A(\mathbf{r},t)$ is deduced as

$$\begin{aligned} A(\mathbf{r},t) &= \sqrt{\mid U^{(r)}(\mathbf{r},t) \mid^2 + \mid U^{(i)}(\mathbf{r},t) \mid^2} \\ &= \sqrt{U(\mathbf{r},t)U^*(\mathbf{r},t)} = |U(\mathbf{r},t)| \end{aligned} \tag{8.43}$$

If we divide $U^{(i)}(\mathbf{r},t)$ by $U^{(r)}(\mathbf{r},t)$, we obtain

$$\frac{U^{(i)}(\mathbf{r},t)}{U^{(r)}(\mathbf{r},t)} = \tan[2\pi\nu_0 t - \psi(\mathbf{r},t)] \tag{8.44}$$

thus, the associated phase factor $\psi(\mathbf{r},t)$ is obtained as

$$\begin{aligned} \psi(\mathbf{r},t) &= \tan^{-1}\frac{U^{(i)}(\mathbf{r},t)}{U^{(r)}(\mathbf{r},t)} - 2\pi\nu_0 t \\ &= \tan^{-1}\left(i\frac{U^*(\mathbf{r},t) - U(\mathbf{r},t)}{U^*(\mathbf{r},t) + U(\mathbf{r},t)}\right) - 2\pi\nu_0 t \end{aligned} \tag{8.45}$$

Here, $A(\mathbf{r},t)$ is independent of ν_0, while $\psi(\mathbf{r},t)$ depends on ν only through the additive term $2\pi\nu_0 t$.

8.3 Emission of Wave-Trains

Light of the spectral line is not strictly monochromatic, but is made up of wave-trains of finite length, of which a large number pass at random time intervals during the time necessary to make an observation. According to the atomic theory, the loss of energy by atoms during emission gives rise to the damping of the wave-train. A very weakly damped wave-train is almost sinusoidal and therefore monochromatic, while a highly damped wave-train corresponds to a simple nonharmonic oscillation and as a result to a non-monochromatic radiation. From an observational point of view, the complex amplitude, $a(t)$, assumes a large number of values that spread over at random. In limited sinusoidal vibration, we have a large number of points distributed at random. In the case of the damped harmonic vibrations, the segments of straight lines that are distributed at random pass through the origin. Consider that all such wave-trains are identical. Let $f(t)$ be the vibration at a given

© 2015 by Taylor & Francis Group, LLC

point at a time, t, due to single wave train. We assume f to be zero for $|t| \geq t_0$, and express it as a Fourier integral

$$f(t) = \int_{-\infty}^{\infty} \widehat{f}(\nu) e^{i2\pi\nu t} d\nu \tag{8.46}$$

where ν is the frequency.

Assuming that an atom emits complex vibrations $f_1(t), f_2(t) \cdots$ at times $t_1, t_2, \cdots$ distributed randomly, the expressions $f_1(t), f_2(t) \cdots$ are the analytic signals associated with real valued waves. The vibration emission at times t_1 is represented by $f_1(t - t_1)$ at time t. If N such wave-trains pass the point during the time required to make an observation, the total light disturbance involved in the observation may be written as

$$\begin{aligned} U(t) &= \sum_{n=1}^{N} f(t_n - t) \\ &= \sum_{n=1}^{N} \int_{-\infty}^{\infty} \widehat{f}(\nu) e^{i2\pi\nu(t_n - t)} d\nu \end{aligned} \tag{8.47}$$

where t_n denote the time of arrival of wave trains and from which we get

$$\begin{aligned} U(t) &= \int_{-\infty}^{\infty} \left[\widehat{f}_1(\nu) e^{-i2\pi\nu t_1} + \widehat{f}_2(\nu) e^{-i2\pi\nu t_2} + \cdots \right] e^{i2\pi\nu t} d\nu \\ &= \int_{-\infty}^{\infty} \widehat{U}(\nu) e^{i2\pi\nu t} d\nu \end{aligned} \tag{8.48}$$

with

$$\widehat{U}(\nu) = \widehat{f}(\nu) \sum_{n=1}^{N} e^{-i2\pi\nu t_n} \tag{8.49}$$

The top panel of Figure 8.5 depicts the time representation of the finite sinusoidal vibration, $U^{(r)}(t)$, and the corresponding spectrum, while the bottom panel shows the time representation of the infinite sinusoidal vibration. The phase of each wave-train is variable and there is no phase relation between the different wave-trains. For each wave-train, according to equation (8.8), we find

$$U(t) = 2 \int_{0}^{\infty} \widehat{U}(\nu) e^{i2\pi\nu t} d\nu \tag{8.50}$$

Equations (8.8 and 8.50) are identical. The complex vibration, $U(t)$, provided by one or the other of these equations represents a succession of wave-trains emitted by an atom. The complex vibration $U(t)$ can represent the vibration emitted at the time t by an extended incoherent source. If the light is quasi-monochromatic, one uses equation (8.40).

Let us consider a detector ([162] and references therein) with a resolution twice as large compared with the coherence time, τ_c; it receives an average of

© 2015 by Taylor & Francis Group, LLC

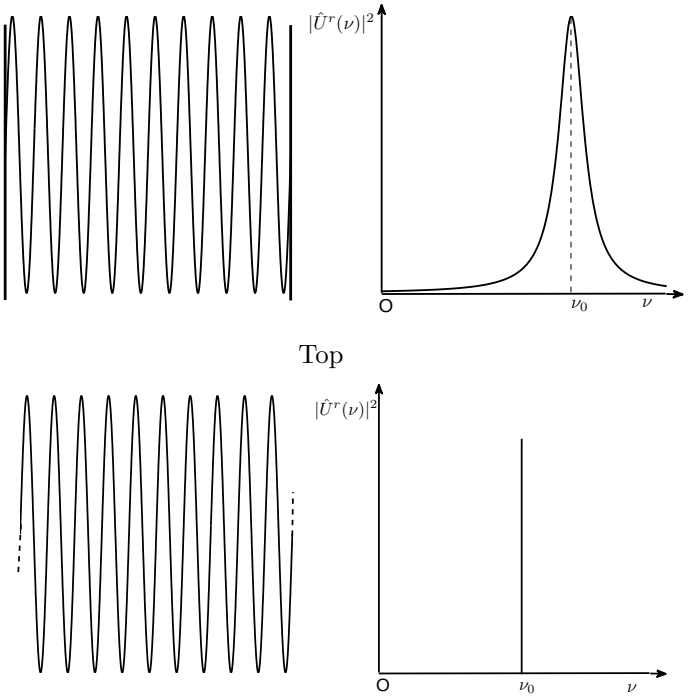


FIGURE 8.5: Top panel left: finite sinusoidal wave-train and right: the corresponding spectrum. Bottom panel left: large number of sinusoidal wave-trains and right: the corresponding spectrum.

the effects produced by the different values of the amplitude, $a(t)$. As stated in Section 1.5, a detector is sensitive to the square of $U^{(r)}(t)$, so that the time average of the intensity tends to a finite value as the averaging interval is increased indefinitely, that is,

$$\begin{aligned} I &= \langle \mid U^{(r)}(t) \mid^2 \rangle \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mid U^{(r)}(t) \mid^2 dt \end{aligned} \tag{8.51}$$

The time necessary to make one observation, T, is very large as compared with the coherence time; therefore, the light intensity averaged over the time interval $2T$ needed to make an observation is given by

$$\begin{aligned} I &= \frac{1}{2T} \int_{-T}^{T} \mid U^{(r)}(t) \mid^2 dt \\ &\simeq \frac{1}{2T} \int_{-\infty}^{\infty} \mid U^{(r)}(t) \mid^2 dt \end{aligned} \tag{8.52}$$

© 2015 by Taylor & Francis Group, LLC

The truncated functions are analyzed by using the Fourier method:

$$U_T^{(r)} = \begin{cases} U^{(r)}(t) & \text{when } |t| \le T \\ 0 & \text{when } |t| > T \end{cases} \tag{8.53}$$

Each of the truncated functions is assumed to be square integrable; thus, in the form of a Fourier integral, we may express it as

$$U_T^{(r)}(t) = \int_{-\infty}^{\infty} \widehat{U}_T(\nu) e^{i2\pi\nu t} d\nu \tag{8.54}$$

Let $U_T^{(i)}$ be the associated function and U_T the corresponding analytic signal, that is,

$$\begin{aligned} U_T(t) &= U_T^{(r)} + iU_T^{(i)}(t) \\ &= 2\int_0^{\infty} \widehat{U}_T(\nu) e^{i2\pi\nu t} d\nu \end{aligned} \tag{8.55}$$

A basic property of the spectral density follows from the Parseval theorem:

$$\begin{aligned} \frac{1}{2T}\int_{-\infty}^{\infty} |\, U_T^{(r)}(t)\,|^2\, dt &= \frac{1}{2T}\int_{-\infty}^{\infty} |\, U_T^{(i)}(t)\,|^2\, dt = \frac{1}{4T}\int_{-\infty}^{\infty} U_T(t)U_T^*(t)dt \\ &= \int_{-\infty}^{\infty} \widehat{G}_T(\nu)d\nu = 2\int_0^{\infty} \widehat{G}_T(\nu)d\nu \end{aligned} \tag{8.56}$$

with $\widehat{G}_T(\nu) = \frac{|\widehat{U}_T(\nu)|^2}{2T}$.

The function $\widehat{G}_T(\nu)$, known as a periodogram, does not tend to a limit, but fluctuates with increasing T. The average of $\widehat{G}_T(\nu)$ taken over the ensemble of the functions $U^{(r)}(t)$ tends to a definite limit as $T \to \infty$. Thus, the smoothed periodogram $\overline{\widehat{G}_T(\nu)}[= \overline{|\widehat{U}_T(\nu)|^2}/(2T)]$, in which the bar denotes the ensemble average, possess a limit:

$$\widehat{G}(\nu) = \lim_{T\to\infty} \overline{\widehat{G}_T(\nu)} = \lim_{T\to\infty} \frac{\overline{|\,\widehat{U}_T(\nu)\,|^2}}{2T}$$

In the limit $T \to \infty$, we find

$$\langle |\, U_T^{(r)}(t)\,|^2 \rangle = \langle |\, U_T^{(i)}(t)\,|^2 \rangle = \frac{1}{2}\langle U_T(t)U_T^*(t)\rangle = 2\int_0^{\infty} \widehat{G}(\nu)d\nu \tag{8.57}$$

The function $\widehat{G}(\nu)$ is called the power spectrum of the random process characterized by the ensemble of the function $U^{(r)}(t)$ and is also referred to as the spectral density of the light vibrations. The term $\widehat{G}(\nu)d\nu$ is proportional to the contribution to the intensity from the frequency range $(\nu, \nu + d\nu)$, in which the truncated function, $U^{(r)}(t)$, is considered to be the light vibration.

© 2015 by Taylor & Francis Group, LLC

8.3.1 Coherence Length and Coherence Time

Thus far we have discussed at the start of this section 8.3 that emitted waves are discharged as wave-trains. As stated in Section 1.6, the coherence of two interfering light beams is linked with the duration and consequently with the length of the wave-trains. This length determines the bandwidth of the radiations emitted by the atoms. If all the wave-trains have the same duration during which $U^{(r)}(t)$ is a simple harmonic of the frequency ν_0, then

$$U^{(r)}(t) = \begin{cases} \cos 2\pi\nu_0 t, & \text{when } |t| \leq \tau_c/2 \\ 0 & \text{when } |t| > \tau_c/2 \end{cases} \tag{8.58}$$

where τ_c is the coherence time of the light.

From equations (8.46 and 8.47) we get

$$U^{(r)}(t) = \int_{-\infty}^{\infty} \widehat{U}^{(r)}(\nu) e^{i2\pi\nu t} d\nu \tag{8.59}$$

the Fourier transform of which is

$$\widehat{U}^{(r)}(\nu) = \int_{-\infty}^{\infty} U^{(r)}(t) e^{-i2\pi\nu t} dt \tag{8.60}$$

The Parseval theorem (see Section 8.2.2) provides us with

$$\begin{aligned} \int_{-\infty}^{\infty} |\, U^{(r)}(t)\,|^2\, dt &= \int_{-\infty}^{\infty} |\, U^{(r)}(\nu)\,|^2\, d\nu \\ &= \int_{-\infty}^{\infty} |\, \widehat{f}^{(r)}(\nu)\,|^2 \sum_{n=1}^{N}\sum_{m=1}^{N} e^{i2\pi\nu(t_n - t)} d\nu \end{aligned} \tag{8.61}$$

We have

$$\begin{aligned} \sum_{n=1}^{N}\sum_{m=1}^{N} e^{2\pi i\nu(t_n - t)} &= N + \sum_{n\neq m} e^{i2\pi\nu(t_m - t_n)} \\ &= N + 2\sum_{n<m} \cos 2\pi\nu(t_m - t_n) \end{aligned} \tag{8.62}$$

where t_n are distributed randomly.

From equations (8.52 and 8.61), the mean intensity, I, is recast as

$$I = \frac{N}{2T} \int_{-\infty}^{\infty} |\, \widehat{f}^{(r)}(\nu)\,|^2\, d\nu \tag{8.63}$$

Equation (8.63) states that the mean intensity is proportional to the integral of the intensities $\widehat{I}(\nu) = |\widehat{f}^{(r)}(\nu)|^2$ for incoherent superposition of the

© 2015 by Taylor & Francis Group, LLC

monochromatic components of which a single wave train is made up. According to equation (8.60), the Fourier transform of $U^{(r)}(t)$ is written as

$$\begin{aligned}\widehat{U}^{(r)}(\nu) &= \frac{1}{2}\int_{-\tau_c/2}^{\tau_c/2}\left[e^{i2\pi\nu_0 t}+e^{-i2\pi\nu_0 t}\right]e^{-i2\pi\nu t}dt \\ &= \frac{\tau_c}{2}\left[\frac{\sin\pi(\nu_0-\nu)\tau_c}{\pi(\nu-\nu_0)\tau_c}+\frac{\sin\pi(\nu_0+\nu)\tau_c}{\pi(\nu+\nu_0)\tau_c}\right] \end{aligned} \tag{8.64}$$

Figure 8.6 depicts the curve representing the vibrations of $\widehat{U}^{(r)}(\nu)$ as a function of ν. When the interfering beams are of equal intensity, τ_c is equal to the normalized root-mean square (RMS) width of the square of the visibility function. From a random sequence of periodic wave-trains, we find that the coherence time, τ_c, and the effective spectral width (bandwidth), $\Delta\nu$, of the resulting disturbance are related by the order of magnitude:

$$\tau_c = \frac{1}{\Delta\nu} \tag{8.65}$$

with

$$\begin{aligned}\Delta\nu &= \frac{1}{(\lambda-\Delta\lambda/2)}-\frac{1}{(\lambda+\Delta\lambda/2)} \\ &= \frac{\Delta\lambda}{\lambda^2-(\Delta\lambda/2)^2}\simeq\frac{\Delta\lambda}{\lambda^2}\end{aligned} \tag{8.66}$$

This result states that the coherence time is inversely proportional to the spectral bandwidth. Here, the fractional bandwidth is assumed to be very small; therefore, the term $(\Delta\lambda/2)^2$ is ignored. With the frequency interval, $\nu_0-\tau_c/2\le\nu<\nu_0+\tau_c/2$, over which the intensity is significant with the first zero corresponding to $\nu_0-\nu=\pm1/\tau_c$, it is clear that

$$\Delta\nu=\nu_0-\nu\sim\frac{1}{\tau_c} \tag{8.67}$$

However, if the light is quasi-monochromatic, i.e., $\mathtt{OA}\gg\mathtt{AB}$ in Figure 8.4, the frequency interval turns out to be $\Delta\nu\ll\nu_0$. Therefore, the spectrum becomes narrow enough to prevent the two terms on the RHS of equation (8.64) from overlapping. According to equation (8.67), the spectral bandwidth is on the order of the inverse of the duration of a wave-train. In the ideal case of light from a real source, the effective frequency range of the Fourier spectrum turns out to be on the order of the reciprocal of the duration of a single wave-train.

Atoms are in random thermal motion relative to the observer; therefore, the observed spectra are changed by an effect, called the Doppler effect (see Section 1.4.4). The emitting atoms are disturbed by their neighbor as well yielding irregular modifications of the wave-trains. The averages of the duration of the wave-train and the frequency range of Fourier spectrum satisfy the

© 2015 by Taylor & Francis Group, LLC

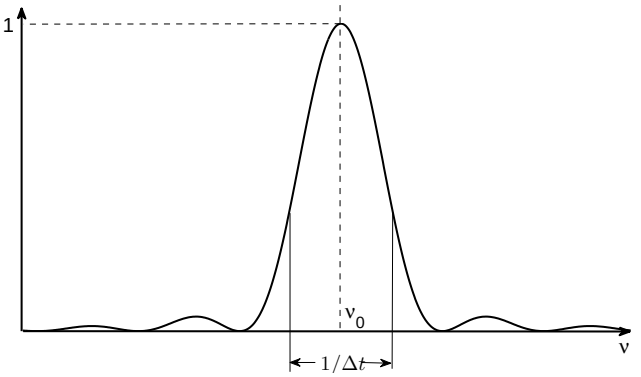


FIGURE 8.6: The function $[\sin \pi(\nu - \nu_0)\tau_c/\pi(\nu - \nu_0)\tau_c]^2$.

reciprocity inequality

$$\tau_c \Delta\nu \geq \frac{1}{4\pi} \tag{8.68}$$

This inequality is analogous to the Heisenberg uncertainty relation in quantum mechanics (see Section 8.1.2.1). Using equation (8.64), equation (8.59) can be recast. Let us set

$$\begin{aligned} \widehat{U}^{(r)}(\nu_0 - \nu) &= \frac{\sin \pi(\nu_0 - \nu)\tau_c}{\pi(\nu_0 - \nu)\tau_c} \\ \widehat{U}^{(r)}(\nu_0 + \nu) &= \frac{\sin \pi(\nu_0 + \nu)\tau_c}{\pi(\nu_0 + \nu)\tau_c} \end{aligned} \tag{8.69}$$

and therefore, we have

$$U^{(r)}(t) = \frac{\tau_c}{2}\int_{-\infty}^{\infty} \widehat{U}^{(r)}(\nu_0 - \nu)e^{i2\pi\nu t}d\nu + \frac{\tau_c}{2}\int_{-\infty}^{\infty} \widehat{U}^{(r)}(\nu_0 + \nu)e^{i2\pi\nu t}d\nu \tag{8.70}$$

Let us simplify equation (8.70) as

$$\begin{aligned} U^{(r)}(t) &= \frac{\tau_c}{2}\int_{-\infty}^{0} \widehat{U}^{(r)}(\nu_0 - \nu)e^{i2\pi\nu t}d\nu + \frac{\tau_c}{2}\int_{0}^{\infty} \widehat{U}^{(r)}(\nu_0 - \nu)e^{i2\pi\nu t}d\nu \\ &+ \frac{\tau_c}{2}\int_{-\infty}^{0} \widehat{U}^{(r)}(\nu_0 + \nu)e^{i2\pi\nu t}d\nu + \frac{\tau_c}{2}\int_{0}^{\infty} \widehat{U}^{(r)}(\nu_0 + \nu)e^{i2\pi\nu t}d\nu \end{aligned} \tag{8.71}$$

which leads to

$$U^{(r)}(t) = \tau_c \int_0^{\infty} \left[\widehat{U}^{(r)}(\nu_0 + \nu) + \widehat{U}^{(r)}(\nu_0 - \nu)\right] \cos[2\pi\nu t]d\nu \tag{8.72}$$

© 2015 by Taylor & Francis Group, LLC

If the spectrum is narrow enough, so that $\widehat{U}^{(r)}(\nu_0+\nu)$ and $\widehat{U}^{(r)}(\nu_0-\nu)$ do not overlap, we can write

$$U^{(r)}(t) \simeq \tau_c \int_0^\infty \widehat{U}^{(r)}(\nu_0-\nu)\cos[2\pi\nu t]d\nu \tag{8.73}$$

The distribution of energy in the spectrum of the vibration, $U^{(r)}(t)$, can be deduced by comparison with equation (8.9):

$$|\,\widehat{U}^{r}(\nu)\,|^2 = (\tau_c)^2\left[\frac{\sin\pi(\nu_0-\nu)\tau_c}{\pi(\nu_0-\nu)\tau_c}\right]^2 \tag{8.74}$$

in which $|\,\widehat{U}^{r}(\nu)\,|^2$ is the spectral distribution of the energy.

The sinusoidal function, $[\sin\pi(\nu-\nu_0)\tau_c/\pi(\nu-\nu_0)\tau_c]^2$, governing the intensity distribution of Fourier component is depicted in Figure 8.6 The value of coherence time, τ_c, is, in general, very small for thermal light sources with respect to the time required for an observation. For a monochromatic thermal source, the value of such a time is on the order of 10^{-8} seconds, while for a laser source, it may be of the order on 10^{-2} seconds. Therefore, the complex amplitude $a(t)$ assumes a large number of values which are distributed at random during the time of one observation. With c as the speed of light in free space, the coherence length, l_c, is defined as

$$l_c = c\tau_c \tag{8.75}$$

$$= \frac{c}{\Delta\nu} = \frac{\lambda_0^2}{\Delta\lambda} \tag{8.76}$$

where $\lambda = c/\nu$ and λ_0 is the mean wavelength.

8.3.2 Damped Harmonic Vibrations

In the case of a damped harmonic vibration, its amplitude decreases with time, eventually going to zero (see Figure 8.7; left panel). The segments of straight lines that are distributed at random pass through the origin[122]. With the damping constant (coherence time), τ_c, the damped vibration is

$$U^{(r)}(t) = e^{-t/\tau_c}\cos 2\pi\nu_0 t \qquad t>0 \tag{8.77}$$

According to equation (8.16), the Fourier transform is

$$\widehat{U}^{(r)}(\nu) = \int_0^\infty e^{-t/\tau_c}\cos 2\pi\nu_0 t e^{-i2\pi\nu t}dt \tag{8.78}$$

We have

$$\widehat{U}^{(r)}(\nu) = \frac{1}{2}\left[\int_0^\infty e^{-t/\tau_c}e^{i2\pi(\nu_0-\nu)t}dt + \int_0^\infty e^{-t/\tau_c}e^{-i2\pi(\nu_0+\nu)t}dt\right] \tag{8.79}$$

© 2015 by Taylor & Francis Group, LLC

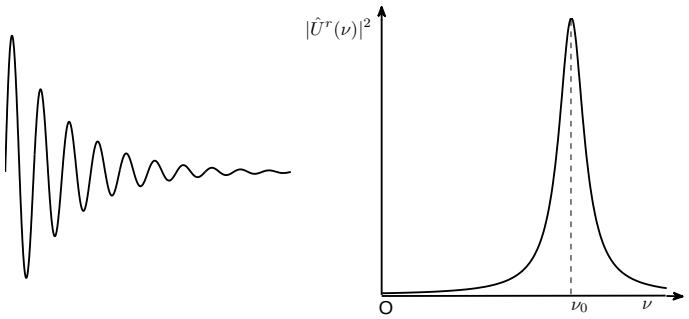


FIGURE 8.7: Left panel: complex amplitudes of a large numbers of decaying wave-trains and right panel: variation of $|\ \widehat{U}^{(r)}(\nu)\ |^2$ as a function of ν.

which follows that

$$\widehat{U}^{(r)}(\nu) = \frac{1}{2}\left[\frac{1}{\tau_c} - i2\pi(\nu_0 - \nu)\right]^{-1} + \frac{1}{2}\left[\frac{1}{\tau_c} + i2\pi(\nu_0 + \nu)\right]^{-1} \tag{8.80}$$

Using equations (8.59 and 8.80), we write

$$U^{(r)}(t) = \frac{1}{2}\left[\int_{-\infty}^{\infty} \frac{e^{i2\pi\nu t}d\nu}{1/\tau_c - i2\pi(\nu_0 - \nu)} + \int_{-\infty}^{\infty} \frac{e^{i2\pi\nu t}d\nu}{1/\tau_c + i2\pi(\nu_0 + \nu)}\right] \tag{8.81}$$

Let us set $\rho e^{i\theta} = 1/\tau_c - i2\pi(\nu_0 - \nu)$ and $\rho' e^{i\theta'} = 1/\tau_c + i2\pi(\nu_0 + \nu)$. We split each integral into two integrals with the limits 0 and ∞. In the first and third integrals we change ν into $-\nu$. Since $\rho' e^{-i\theta'} = 1/\tau_c - i2\pi(\nu_0 + \nu)$ and $\rho e^{-i\theta} = 1/\tau_c + i2\pi(\nu_0 - \nu)$, we get

$$\begin{aligned} U^{(r)}(t) &= \int_0^{\infty} \left[\frac{\cos(2\pi\nu t - \theta')}{\rho'} + \frac{\cos(2\pi\nu t - \theta)}{\rho}\right] d\nu \\ &\simeq \int_0^{\infty} \frac{\cos(2\pi\nu t - \theta)}{\rho} d\nu \end{aligned} \tag{8.82}$$

in which the first term on the RHS of equation (8.82) turns out to be almost negligible in the case of the narrow spectrum in comparison to the second term in the neighborhood of $\nu = \nu_0$.

The spectral distribution of the energy which is displayed in Figure 8.7 (right panel) is given by

$$|\ \widehat{U}^{(r)}(\nu)\ |^2 = \frac{1}{\rho^2} = \left[\frac{1}{\tau_c^2} + 4\pi^2(\nu_0 - \nu)^2\right]^{-1} \tag{8.83}$$

© 2015 by Taylor & Francis Group, LLC

Let us characterize the spectral bandwidth, $\Delta\nu$, by the width corresponding to one half of the maximum ordinate, τ_c^2:

$$\left[\frac{1}{\tau_c^2} + 4\pi^2(\nu_0 - \nu)^2\right]^{-1} = \frac{\tau_c^2}{2}$$

from which we get

$$(\nu_0 - \nu)^2 = \left(\frac{\Delta\nu}{2}\right)^2 = \frac{1}{4\pi^2\tau_c^2}$$

thus,

$$\Delta\nu = \frac{1}{\pi\tau_c} \tag{8.84}$$

The coherence length is given by

$$l_c = \frac{c}{\pi\Delta\nu} = \frac{\lambda_0^2}{\pi\Delta\lambda} \tag{8.85}$$

8.4 Polarization of Plane Monochromatic Waves

Polarization structure of the waves is an important characteristic of the nature of light. Based on the work on the propagation of light through crystal, Huygens suggested that light was not a scalar quantity. As stated earlier in Section 2.6.1, the electric and magnetic field strength of an electromagnetic wave oscillates perpendicularly against the travel direction of the wave. In three dimensions, if the wave propagates in the x direction, the electrical field strength can oscillate both in y and z directions or vary between these two directions in an organized or an unorganized manner. The electromagnetic wave is said to have different modes of polarization.

Fresnel in 1823 derived the reflection and transmission formulae for a plane wave that is incident on a static and plane interface between two dielectric isotropic media. Light can be polarized under natural conditions if the incident light strikes a surface at an angle equal to the polarizing angle of that media. The modified incident polarization caused by the reflection of a mirror is characterized by two parameters:

- the ratio between the reflection coefficients of the electric vector components which are perpendicular and parallel to the plane of incidence, known as s and p components, respectively, and
- the relative phase-shift between these electric vibrations.

© 2015 by Taylor & Francis Group, LLC

The polarization of the wave characterizes how the direction of the electric field vector varies at a given point in space as a function of time. If the direction of vibration remains the same with time, the wave is linearly polarized or plane polarized in that direction. If the direction of vibration rotates at the same frequency as the wave, the wave is said to be circularly polarized. Intermediate states are called partially polarized. The amount of order is specified by the degree of polarization.

Let us examine the propagation of plane electromagnetic waves in free space. In classical mechanics if we equate an optical medium to an isotropic elastic medium[124], it should, in principle, support three independent oscillations, that is, $V_x(\mathbf{r},t)$, $V_y(\mathbf{r},t)$, and $V_z(\mathbf{r},t)$, so that equation (1.16) can be expressed as

$$\left(\boldsymbol{\nabla}^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) V_j(\mathbf{r},t) = 0 \tag{8.86}$$

where v is the velocity of propagation of the oscillation, $j = x, y, z$, $\mathbf{r}$ the position vector.

Maxwell opined that in free space the transverse component would be present. In a Cartesian system, the components, $V_x(\mathbf{r},t)$ and $V_y(\mathbf{r},t)$ are the transverse components, and $V_z(\mathbf{r},t)$ is the longitudinal component when the propagation is in the z direction. Hence, according to equation (8.86), the optical field components are

$$V_x(\mathbf{r},t) = V_{0x}\cos(\boldsymbol{\kappa}\cdot\mathbf{r} - \omega t + \psi_1) \tag{8.87}$$
$$V_y(\mathbf{r},t) = V_{0y}\cos(\boldsymbol{\kappa}\cdot\mathbf{r} - \omega t + \psi_2) \tag{8.88}$$
$$V_z(\mathbf{r},t) = V_{0z}\cos(\boldsymbol{\kappa}\cdot\mathbf{r} - \omega t + \psi_3) \tag{8.89}$$

in which V_{0j} are the maximum amplitudes and $\psi_{i(=1,2,3)}$ the arbitrary phases.

Fresnel–Argo carried out a series of investigations on the Young's interference experiment using polarized light and concluded that the longitudinal component, $V_z(\mathbf{r},t)$, did not exist. So, the propagation of light in z direction would be

$$V_x(z,t) = V_{0x}\cos(\kappa z - \omega t + \psi_1) \tag{8.90}$$
$$V_y(z,t) = V_{0y}\cos(\kappa z - \omega t + \psi_2) \tag{8.91}$$

A polarized light can be extinguished by an ideal polarizer, that is, a polarizer with an intensity transmittance of one for its principal state and an intensity transmittance of zero for the orthogonal state. For polychromatic light (see Section 8.2), the polarization ellipses associated with each spectral component have identical ellipticity which is the ratio of the minor axis to the major axis of the corresponding electric field polarization ellipse, orientation, and helicity.

© 2015 by Taylor & Francis Group, LLC

8.4.1 Instantaneous Optical Field and Polarization Ellipse

The behavior of the electric field vector, **E**, at a particular time in space and its changes from point to point of the field with time are captured by the state of polarization of a wave. This state of polarization enters many physical phenomena, such as reflection and transmission. The amount of reflected and transmitted light depends on the state of polarization. As stated in Section 8.4, light consists of two transverse components which are perpendicular to each other and can be chosen for convenience to be propagating in the z direction. The waves are said to be instantaneous with equal to the time duration for the wave to go through one complete cycle equal to 10^{-15} s at optical frequencies.

The plane waves described by equations (2.265 and 2.266) are transverse implying **E** and **B** oscillate in a plane perpendicular to the wave number vector, $\boldsymbol{\kappa}$. Let us assume an electric field of a plane electromagnetic uniform light beam propagating in the positive direction of the z axis of right-handed Cartesian coordinate system (x, y, z) at a point **r** in space. The electric fields associated with the waves are written in the form:

$$\mathbf{E}_1(\mathbf{r}, t) = \hat{\mathbf{x}}_1 a_x \cos(\kappa z - \omega t + \psi_1) \tag{8.92}$$

$$\mathbf{E}_2(\mathbf{r}, t) = \hat{\mathbf{x}}_2 a_y \cos(\kappa z - \omega t + \psi_2) \tag{8.93}$$

which fulfill the wave equation. Here a_x and a_y are the instantaneous amplitudes of the two orthogonal components, ψ_1 and ψ_2 are the respective instantaneous phases at a fixed point in space as a function of time, and $\hat{\mathbf{x}}_{j(=1,2)}$ are the unit vectors. The resultant of these waves is $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$ which can be written as

$$\mathbf{E} = \hat{\mathbf{x}} a \cos(\kappa z - \omega t + \psi) \tag{8.94}$$

where $a[= \sqrt{a_x^2 + a_y^2 = 2a_x a_y \cos(\psi_1 - \psi_2)}]$ represents the amplitude of the wave.

Equation (8.94) states that the resultant is a linearly polarized wave with its electric vector oscillating along the same axis. Let us consider the superposition of two linearly polarized electromagnetic waves, which propagate along the z axis, but with their electric vectors oscillating along two mutually perpendicular directions. Therefore,

$$\mathbf{E}_1 = \hat{\mathbf{x}} a_x \cos(\kappa z - \omega t) \tag{8.95}$$

$$\mathbf{E}_2 = \hat{\mathbf{x}} a_y \cos(\kappa z - \omega t + \psi) \tag{8.96}$$

Any polarization state may be described as the resultant of two orthogonal linear polarization states. Let us consider the time variation of the resultant electric field at an arbitrary plane perpendicular to the z axis which we may, without any loss of generality, assume to be $z = 0$. Let an electric field for the plane wave be propagate in the direction of the $\pm z$ axis in free space:

$$\begin{aligned} \mathbf{E}_{\pm}(z, t) &= \mathbf{E}_{x\pm}(z, t)\hat{\mathbf{x}}_1 + \mathbf{E}_{y\pm}(z, t)\hat{\mathbf{x}}_2 \\ &= a_x \cos(\kappa z \mp \omega t + \psi_1)\hat{\mathbf{x}}_1 + a_y \cos(\kappa z \mp \omega t + \psi_2)\hat{\mathbf{x}}_2 \end{aligned} \tag{8.97}$$

© 2015 by Taylor & Francis Group, LLC

in which $\mathbf{E}_x$, and $\mathbf{E}_y$ are the electric vectors, along the x and y axes, respectively.

The most general form of polarization is elliptical polarization in which the end points of the instantaneous electric vectors lie on an ellipse. The other states of polarization are its corollaries. As the monochromatic wave propagates through space in a fixed plane perpendicular to the direction of light, the end point of the electric vector at a fixed point traces out an ellipse. The shape of the ellipse changes continuously. When the ellipse maintains a constant orientation, ellipticity, and sense in the ellipse, the wave is known to be completely polarized at that point.

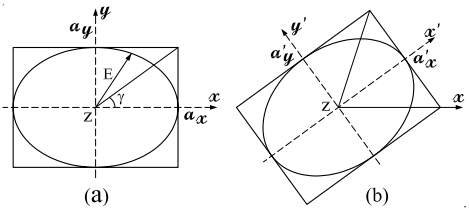


FIGURE 8.8: Description of polarization ellipse (a) in terms of $x,\ y$ and (b) in terms of $x',\ y'$ coordinates.

Let us consider that the transverse components are represented by

$$E_x(z,t) = a_x \cos(\psi_1 - \omega t) \tag{8.98}$$

$$E_y(z,t) = a_y \cos(\psi_2 - \omega t) \tag{8.99}$$

These signals (equations 8.98 and 8.99) fluctuate slowly in comparison with the cosine term at optical frequencies. As the field propagates, $E_x(z,t)$ and $E_y(z,t)$ give rise to a resultant vector describing a locus of points in space. The polarization ellipse is circumscribed in a rectangle (see Figure 8.8), the sides of which are parallel to the x and y axes in which the angle between the diagonal and the x axis is γ (see Figure 8.8a). The propagation is in the z direction. Since the field is transverse, the x and y components of this electric field are different from zero. The expression for the trajectory parametrized by these equations is obtained by eliminating ωt:

$$\begin{aligned}\left(\frac{E_x}{a_x}\right)^2 + \left(\frac{E_y}{a_y}\right)^2 &= \cos^2(\psi_1 - \omega t) + \cos^2(\psi_2 - \omega t) \\ &= \sin^2(\psi_1 - \psi_2) + 2\frac{E_x}{a_x}\frac{E_y}{a_y}\cos(\psi_1 - \psi_2) \\ &= \sin^2\delta + 2\frac{E_x}{a_x}\frac{E_y}{a_y}\cos\delta \end{aligned} \tag{8.100}$$

© 2015 by Taylor & Francis Group, LLC

where $\delta = \psi_1 - \psi_2$, $-\pi < \delta \leq \pi$, is the phase difference between the x and y vibrations.

By rearranging equation (8.100), we get

$$\left(\frac{E_x}{a_x}\right)^2 + \left(\frac{E_y}{a_y}\right)^2 - 2\frac{E_x}{a_x}\frac{E_y}{a_y}\cos\delta = \sin^2\delta \tag{8.101}$$

Equation (8.101) describes an ellipse and shows that at any instant of time the locus of points described by the optical field as it propagates is an ellipse. This behavior is spoken of as optical polarization and is referred to as the polarization ellipse of the electric field for a monochromatic light in which the amplitudes and phases are constant. This equation can be written as

$$a_x^2 E_y^2 - (2a_x a_y E_x \cos\delta)E_y + a_y^2(E_x^2 - a_x^2 \sin^2\delta) = 0 \tag{8.102}$$

The solution of quadratic equation (8.102) is

$$E_y = \frac{a_y E_x \cos\delta}{a_x} \pm \frac{a_y \sin\delta}{a_x}\sqrt{a_x^2 - E_x^2} \tag{8.103}$$

Let us differentiate equation (8.103) and set $E_y' = dE_y/dE_x = 0$. We get

$$E_x = \pm a_x \cos\delta \tag{8.104}$$

Let us put equation (8.104) into equation (8.103); we get

$$\begin{aligned} E_y &= \frac{a_y a_x \cos^2\delta}{a_x} \pm \frac{a_y \sin\delta}{a_x}\sqrt{a_x^2 - a_x^2\cos^2\delta} \\ &= a_y\cos^2\delta \pm a_y \sin^2\delta = \pm a_y \end{aligned} \tag{8.105}$$

Similarly, by considering equation (8.103), in which the slope is $E_y' = \infty$ on the side of the right angle:

$$E_x = \pm a_x; \qquad E_y = \pm a_y \cos\delta \tag{8.106}$$

Equations (8.105 and 8.106) show that the maximum length of the sides of the ellipse are $E_x = \pm a_x$ and $E_y = \pm a_y$. The ellipse is tangent to the sides of the rectangle at $(\pm a_x, \pm a_y \cos\delta)$ and $(\pm a_x \cos\delta, \pm a_y)$.

When $\delta = \pi/2$, equation (8.101) reduces to

$$\left(\frac{E_x}{a_x}\right)^2 + \left(\frac{E_y}{a_y}\right)^2 = 1 \tag{8.107}$$

which means the reference axes coincide with the major and minor axes of the ellipse. In rest of the cases, the reference axes do not correspond to the major and minor axes of the ellipse. This state is an elliptic polarization. The cross-term $E_x E_y$ of equation (8.101) implies that the polarization ellipse of the electric field rotates through an angle. Figure 8.9 depicts the different

© 2015 by Taylor & Francis Group, LLC

Lissajous representations of polarization ellipses for different phase angles. For $\psi = n\pi$, in which $n = 0, \pm 1, \pm 2, \cdots$, **E** is linearly polarized. Such an angle depends on the relative values of a_x and a_y. This implies that the ellipse reduces to a segment of straight lines. When θ turns out to be $n\pi/2$, where $n = \pm 1, \pm 3, \cdots$, **E** becomes circularly polarized, implying that the ellipse degenerates to a circle.

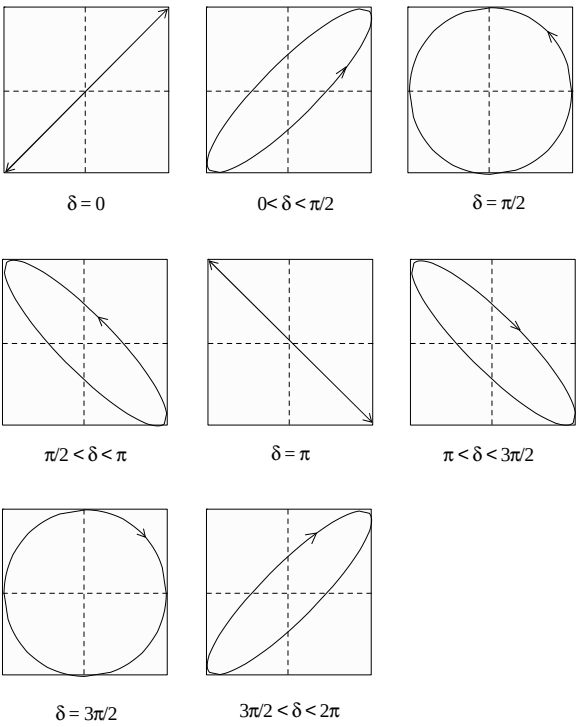


FIGURE 8.9: Lissajous representations of polarization ellipses for various values of the phase difference δ.

For monochromatic radiation, the amplitudes and phases are constant for all time, so equation (8.101) may be expressed as

$$\frac{E_x^2(t)}{a_x^2} + \frac{E_y^2(t)}{a_y^2} - 2\frac{E_x(t)E_y(t)}{a_x\, a_y}\cos\delta = \sin^2\delta \tag{8.108}$$

For $\delta = 0$, we have linear polarization; the field components E_x and E_y are in phase and the polarization ellipse reduces to a straight line. For linearly polarized waves, the amplitudes are constant with respect to magnitude and direction, and both E_x and E_y have the same phase. Let us assume $\psi_1 = \psi_2$;

© 2015 by Taylor & Francis Group, LLC

hence, equation (8.108) takes the form

$$\left(\frac{E_x(t)}{a_x} - \frac{E_y(t)}{a_y}\right)^2 = 0 \tag{8.109}$$

or

$$a_y E_x(t) - a_x E_y(t) = 0 \tag{8.110}$$

which is an expression for a straight line in the E_x, E_y plane.

Since the amplitudes a_x and a_y are nonnegative, this line lies in the first and third quadrants; thus, $E(z,t)$ oscillates along this line. Similarly, if $\psi_1 - \psi_2 = \pi$, equation (8.108) translates into

$$a_y E_x(t) + a_x E_y(t) = 0 \tag{8.111}$$

which lies in the second and fourth quadrants.

If the magnitudes of a_1 and a_2 are equal, but exhibit a phase difference of $\delta = \pm\pi/2$, the situation is known to be a state of circular polarization. The relative amplitude of the components E_x and E_y of the field is described by an auxiliary angle for polarization ellipse, γ, which is given by

$$\tan\gamma = \frac{a_y}{a_x} \qquad 0 < \gamma < \frac{\pi}{2} \tag{8.112}$$

The shape and the orientation of the polarization ellipse and the senses in which the said ellipse is traced out are determined by the angles γ and δ. The size of the ellipse depends on the amplitude of the electric field.

8.4.2 Stokes Parameters

A major advancement to the study of polarization came from the introduction of four measurable quantities for elucidating the properties of light. The state of polarization of a beam of light is, in general, described by four measurable quantities that specify the intensity of the field, the degree of polarization, the plane of polarization, and the ellipticity of the radiation at each point and in any given direction. These quantities at optical frequencies, known as Stokes parameters ([82], [112]) were introduced by George Gabriel Stokes (1819–1903). Stokes vector, S, describes light polarization as

$$S = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \tag{8.113}$$

in which S_0 represents the total intensity of light, S_1 the intensity difference between horizontal and vertical linearly polarized components, S_2 the intensity difference between linearly polarized components oriented at $\pm 45°$, and S_3 the

© 2015 by Taylor & Francis Group, LLC

intensity difference between right-handed and left-handed circular components contained within the beam.

To note, mathematically S is not a vector; however, it is referred to as a vector. To be precise, equation (8.113) should be called the Stokes column matrix. In the case of a quasi-monochromatic approximation (see Section 8.2.4), Stokes parameters measure the average over an ensemble or over time of fields; therefore, equation (8.108) can be recast as

$$\frac{\langle E_x^2(t)\rangle}{a_x^2} + \frac{\langle E_y^2(t)\rangle}{a_y^2} - 2\frac{\langle E_x(t)E_y(t)\rangle}{a_x\, a_y}\cos\delta = \sin^2\delta \tag{8.114}$$

where

$$\langle E_i(t)E_j(t)\rangle = \lim_{T\to\infty}\frac{1}{T}\int_0^T E_i(t)E_j(t)dt \qquad i,j = x,y \tag{8.115}$$

At a particular point in space, the four instantaneous Stokes parameters are defined by the sample realizations at time t of the following combinations of the complex analytical signal $E_x(t)$ and $E_y(t)$ associated with the real-valued components of the electric vector in mutually orthogonal directions perpendicular to $\hat{\mathbf{x}}$. In terms of complex amplitudes, the signals are expressed as

$$E_x(t) = \mathcal{E}_x e^{-i\omega t} \tag{8.116}$$
$$E_y(t) = \mathcal{E}_y e^{-i\omega t} \tag{8.117}$$

in which $\mathcal{E}_x = a_x e^{i\psi_1}$ and $\mathcal{E}_y = a_y e^{i\psi_2}$ are complex amplitudes.

Stokes parameters for a plane wave are now obtained from the formulae:

$$\begin{aligned} S_0 &= E_x E_x^* + E_y E_y^* \\ S_1 &= E_x E_x^* - E_y E_y^* \\ S_2 &= E_x E_y^* + E_y E_x^* \\ S_3 &= i\left(E_x E_y^* - E_y E_x^*\right) \end{aligned} \tag{8.118}$$

in which E^* represents the complex conjugate of E.

By using Stokes parameters[172], the state of polarization of a beam of light is described as

$$\begin{aligned} S_0 &= \mid E_x(t)\mid^2 + \mid E_y(t)\mid^2 = \langle a_x^2\rangle + \langle a_y^2\rangle \\ S_1 &= \mid E_x(t)\mid^2 - \mid E_y(t)\mid^2 = \langle a_x^2\rangle - \langle a_y^2\rangle \\ S_2 &= \frac{1}{2}\left(\mid E_x(t)+E_y(t)\mid^2 - \mid E_x(t)-E_y(t)\mid^2\right) \\ &= E_x(t)E_y^*(t) + E_x^*(t)E_y(t) = 2\mid E_x(t)\mid\mid E_y(t)\mid\cos\delta = 2\langle a_x\, a_y\rangle\cos\delta \\ S_3 &= \frac{1}{2}\left(\mid E_x(t)+iE_y(t)\mid^2 - \mid E_x(t)-iE_y(t)\mid^2\right) \\ &= i\left(-E_x^*(t)E_y(t) + E_x(t)E_y^*(t)\right) = 2\mid E_x(t)\mid\mid E_y(t)\mid\sin\delta \\ &= 2\langle a_x\, a_y\rangle\sin\delta \end{aligned} \tag{8.119}$$

© 2015 by Taylor & Francis Group, LLC

In the case of a monochromatic wave, a_1, a_2, and δ are independent of time; thus, the aforementioned equations reduce to

$$\begin{aligned} S_0 &= a_x^2 + a_y^2 \\ S_1 &= a_x^2 - a_y^2 \\ S_2 &= 2a_x a_y \cos\delta \\ S_3 &= 2a_x a_y \sin\delta \end{aligned} \tag{8.120}$$

These four equations (8.120) are Stokes polarization parameters for a plane wave having the physical dimensions of time-average power per unit area. These parameters are not independent, since

$$S_0^2 = S_1^2 + S_2^2 + S_3^2 \tag{8.121}$$

Let us consider a partially polarized light, in which case the relations given by equation (8.120) continue to be valid for very short time intervals, since the amplitudes and phases fluctuate slowly. Using the Schwarz inequality, we may envisage that for any state of polarized light, Stokes parameters satisfy the relation:

$$S_0^2 \geq S_1^2 + S_2^2 + S_3^2 \tag{8.122}$$

To note, the equality sign applies for the completely polarized light, while the inequality sign is for the partially polarized light or unpolarized light. If the light is emitted by a tungsten lamp, the electrical field strength oscillates in all directions and randomly in the yz plane over time with the same probability. When there is a definite relationship between the time and the direction of the electric field strength, the light is unpolarized. It is a mixture of linearly polarized components in all possible directions, as well as all possible circular and elliptical polarizations. The unpolarized light originates in thermal sources. We may assume that 50% of the intensity of an incoming unpolarized wave falls along the y axis, while the other 50% falls along the z axis. The partially polarized light can be thought of as a mixture of polarized and unpolarized light.

Let us define a vector a_0 whose amplitude is $[a_x^2 + a_y^2]^{1/2}$ and which makes an angle θ with the positive x axis such that $a_x = a_0 \cos\theta$ and $a_x = a_0 \sin\theta$. Here, θ is the position angle of polarization. Both a_x and a_y may take a positive or negative sign defining the quadrant in which θ lies. In terms of the positive angle of polarization and the phase difference δ, Stokes parameters are deduced as

$$\begin{aligned} S_0 &= a_0^2 \\ S_1 &= S_0 \cos 2\theta \\ S_2 &= S_0 \sin 2\theta \cos\delta \\ S_3 &= S_0 \sin 2\theta \sin\delta \end{aligned} \tag{8.123}$$

In the case of $\delta = 0$, the vibration lies in a plane which makes an angle θ

© 2015 by Taylor & Francis Group, LLC

with the x-y plane. When $\delta = \pm\pi$, the light turns out to be plane polarized. If the magnitudes of a_1 and a_2 are equal, but the phase difference of $\delta = \pm\pi/2$, the major and minor axes of the ellipse traced by the instantaneous electric vectors coincide with the x and y axes. Such a situation is said to be circularly polarized. When the δ has any value other than the aforementioned values, the resultant electric vector traces an ellipse in the x-y plane with the major axis arbitrarily inclined to the x axis:

$$\begin{aligned} \tan 2\beta &= \tan 2\theta \cos\delta \\ \sin 2\chi &= \sin 2\theta \sin\delta \end{aligned} \tag{8.124}$$

with

$$\tan 2\beta = \frac{S_2}{S_1} \quad \text{and} \quad \sin 2\chi = \frac{S_3}{S_0} \tag{8.125}$$

The angles β and χ specify the orientation of the ellipse and characterize the ellipticity and sense in which the polarization ellipse is being described, respectively.

8.4.2.1 Poincaré Sphere

The state of polarization is indicated by the location of a point P on the surface of a sphere of radius S_0 in the subspace of S_1, S_2, and S_3, known as a Poincaré sphere, which was proposed by Henri Poincaré (1854−1912). This method provides a convenient graphical tool for utilizing changes in the state of polarization[83]. Each point on the sphere represents a definite state of polarization. For partially polarized light, the radius is smaller than S_0 and it becomes zero in the case of unpolarized light. There is one to one correspondence between the state of polarization and the points on this sphere. The parameter S_0 is proportional to the intensity of the wave. The other parameters, S_1, S_2, and S_3 are related to the orientation angle β $(0 \leq \beta < \pi)$ and the ellipticity angle χ $(-\pi/4 \leq \chi \leq \pi/4)$ of the ellipse. (S_1, S_2, S_3) are regarded as the Cartesian co-ordinate of a point P of radius S_0, such that 2β and 2χ are the spherical angular co-ordinates of the point, i.e., the longitude and latitude, respectively. The lines of longitude and latitude represent the equi-azimuth and equi-ellipticity contours, respectively.

The three co-ordinates (S_1, S_2, S_3) of a point (S_0, χ, β) on the Poincaré sphere (see Figure 8.10) are given by

$$\begin{aligned} S_1 &= S_0 \cos 2\chi \cos 2\beta \\ S_2 &= S_0 \cos 2\chi \sin 2\beta \\ S_3 &= S_0 \sin 2\chi \end{aligned} \tag{8.126}$$

Stokes parameters described by equation (8.126) are identical in form to the equations relating Cartesian coordinates to spherical coordinates (see Appendix B.1.3). Comparing equations (8.126 and B.12), we find that the angles

© 2015 by Taylor & Francis Group, LLC

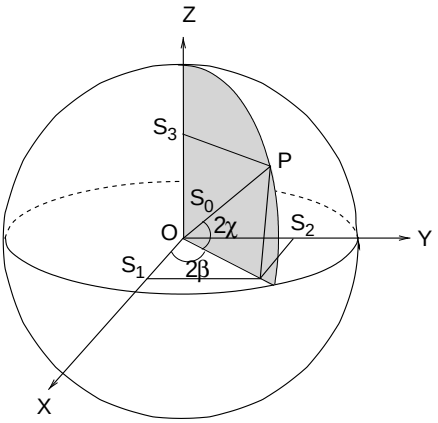


FIGURE 8.10: Poincaré representation of polarized light on a sphere.

are identical if the angles are related by $\theta = 90^\circ - 2\chi$ and $\phi = 2\beta$. Let us arrange this expression in the form of a Stokes vector; thus,

$$S = S_0 \begin{bmatrix} 1 \\ \cos 2\chi \cos 2\beta \\ \cos 2\chi \sin 2\beta \\ \sin 2\chi \end{bmatrix} \tag{8.127}$$

8.4.2.2 Degree of Polarization

Stokes parameters enable us to describe the degree of polarization, $\mathcal{P}$, for any state of polarization. By definition, $\mathcal{P}$ is given by

$$\mathcal{P} = \frac{I_{pol}}{I_{tot}} = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \qquad 0 \leq \mathcal{P} \leq 1 \tag{8.128}$$

where I_{pol} is the intensity of the sum of the polarization components and I_{tot} the total intensity of the beam.

Mathematically, on the Poincaré sphere, $\mathcal{P}$ represents the distance of a normalized Stokes vector's last three components from the origin. The surface of the unit Poincaré sphere corresponds to $\mathcal{P} = 1$ and represents all totally polarized states[108]. If a polarized wave interacts with a random medium, the polarization state of the back-scattered wave may get changed. Therefore, depolarization is associated with a reduction in the polarization of incident states. The degree of depolarization, $\mathcal{P}'$ is defined as

$$\mathcal{P}' = 1 - \mathcal{P} \tag{8.129}$$

The other parameters of the degree of polarization are defined as

© 2015 by Taylor & Francis Group, LLC

1. Degree of linear polarization: The properties of linearly polarized light are (i) the direction of polarization is perpendicular to the direction of propagation and (ii) the electric field vector oscillates in the plane of polarization. The degree of linear polarization is expressed as

$$\mathcal{P}_L = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \tag{8.130}$$

2. Degree of circular polarization: In circular polarization, the electric field maps out a circular pattern. The degree of circular polarization is given by

$$\mathcal{P}_C = \frac{S_3}{S_0} \tag{8.131}$$

3. Degree of polarization uniformity:

$$\mathcal{P}_U = \sqrt{S_1^2 + S_2^2 + S_3^2} \tag{8.132}$$

4. Linear polarization ratio:

$$\mu_L = \frac{S_0 - S_1}{S_0 + S_1} \tag{8.133}$$

5. Circular polarization ratio:

$$\mu_C = \frac{S_0 - S_3}{S_0 + S_3} \tag{8.134}$$

The state of partially elliptically polarized light is described in terms of combination of two independent components:

- natural unpolarized light of intensity $S_0(1 - \mathcal{P}_e)$ and
- fully elliptically polarized light of intensity $S_0\mathcal{P}_e$, in which $\mathcal{P}_e = S_0\mathcal{P}/S_0$ is the degree of polarization, $S_0\mathcal{P}$ the polarized part of the intensity.

Therefore, Stokes parameters are modified to

$$\begin{aligned} S_1 &= S_0\mathcal{P}_e \cos 2\chi \cos 2\beta \\ S_2 &= S_0\mathcal{P}_e \sin 2\chi \cos 2\beta \\ S_3 &= S_0\mathcal{P}_e \sin 2\beta \end{aligned} \tag{8.135}$$

where

$$\mathcal{P}_e = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \qquad 0 \le \mathcal{P}_e \le 1$$

and the factor $\mathcal{P}_e \cos 2\beta = \mathcal{P}$ is known as the degree of polarization.

© 2015 by Taylor & Francis Group, LLC

Thus, the subequations (8.135) translate into

$$\begin{aligned} S_1 &= S_0 \mathcal{P} \cos 2\chi \\ S_2 &= S_0 \mathcal{P} \sin 2\chi \\ S_3 &= S_0 \mathcal{P}_v \end{aligned} \tag{8.136}$$

in which $\mathcal{P}_v = \mathcal{P}_e \sin 2\beta$ is the degree of ellipticity, positive for right-handed elliptical polarization and negative for left-handed polarization.

For a beam of light, the quantities, S, $(S_1^2 + S_2^2)$, and S_3 are invariant under a rotation of the reference coordinate system. For a celestial object, the equatorial coordinate system provides a system of reference with the positive x axis in the direction of north defined by the meridian through the object and the positive y axis in the direction of east, the direction of positive z axis being along the line of sight towards the observer. From equation (8.125), we may recast the orientation angle, β, as well as the ellipticity angle, χ, as

$$\begin{aligned} \beta &= \frac{1}{2} \tan^{-1} \frac{S_2}{S_1} \\ \chi &= \frac{1}{2} \sin^{-1} \frac{S_3}{S_0} \end{aligned} \tag{8.137}$$

Since $\delta = 0$ for plane polarized light, equation (8.124) provides $\chi = \theta$, the positive angle of polarization. The ellipticity which is the ratio of the minor axis to the major axis of the corresponding electric field polarization ellipse, e, is given by

$$e = \frac{S_3}{S_0 + \sqrt{S_1^2 + S_2^2}} \tag{8.138}$$

It varies from 0 for linearly polarized light to 1 for circularly polarized light. The polarization ellipse is alternatively described by its eccentricity, which is represented by

$$\epsilon = \sqrt{1 - e^2} \tag{8.139}$$

The eccentricity is zero for circularly polarized light, increases as the ellipse becomes thinner, and becomes one for linearly polarized light.

8.4.3 Measurements of Stokes Parameters

When an optical beam interacts with matter, its polarization state is changed due to (i) change in amplitude, (ii) change in phase, (iii) change in direction of the orthogonal field components, and (iv) transferring energy from polarized state to the unpolarized state. If the energy in polarized states goes to unpolarized state, the element is known as depolarizer. These effects can be understood by writing the transverse field components for a plane wave:

$$\begin{aligned} E_x(z,t) &= a_x \cos(\kappa z - \omega t + \psi_1) \\ E_y(z,t) &= a_y \cos(\kappa z - \omega t + \psi_2) \end{aligned} \tag{8.140}$$

© 2015 by Taylor & Francis Group, LLC

Equation (8.140) can be changed by varying the amplitudes, a_x and a_y, or the phase, ψ_x or ψ_y, and the direction of $E_x(z,t)$ and $E_y(z,t)$. In what follows, we enumerate the corresponding optical devices for changing the afore-mentioned characteristics.

8.4.3.1 Polarizer

An optical element which changes the orthogonal amplitudes unequally is called a polarizer, or diattenuator. A polarizer is a diattenuating optical element designed to transmit light in a specified polarization state independent of the incident polarization state. The transmission of one of the eigen-polarizations is very nearly zero. The eigen-polarization is a polarization state transmitted unaltered by a polarization element except for a change of amplitude and phase. The eigen-polarizations are the eigen-vectors of the corresponding Jones or Mueller matrix.

The diattenuation is the property of an optical element or system whereby the intensity transmittance of the exiting beam depends on the polarization state of the incident beam. The intensity transmittance is a maximum, P_{max}, for one incident state and a minimum, P_{min}, for the orthogonal state. The diattenuation is defined as $(P_{max} - P_{min})/(P_{max} + P_{min})$. The diattenuator is an element, which can be linear, circular, or in general, elliptical, depending on the type of polarization that emerges. There is a large variety of polarizers with their respective advantages and disadvantages. Dichroism and birefringence are two properties of materials that have been employed to produce optical elements with specific polarization behavior. The former is the material property of displaying diattenuation during propagation. For each direction of propagation, dichroic media have two modes of propagation with different absorption coefficients. The size distribution of grains responsible for producing dichroic scattering may be obtained by studying the wavelength dependence of multiwavelengths. The latter is a material property by which certain materials have two different refractive indices for two orthogonal polarization components. This property is observed mostly in an anisotropic crystal structure having an axis of symmetry, known as a principal axis, such as quartz and calcite. For each propagation direction within a birefringent medium, there are two modes of propagation with different refractive indices, n_1 and n_2. The birefringence, Δn is equal to $\mid n_1 - n_2 \mid$.

The linear polarizers can be divided into absorptive polarizes and beam-splitting polarizers. A linear polarizer is a device which, when placed in an incident unpolarized beam, produces a beam of light whose electric field vector is oscillating primarily in one plane, with only a small component in the perpendicular plane[104]. As stated at the begining of Section 8.4, when light reflects at an angle from an interface between two transparent materials, the reflectivity is different for light polarized in the plane of incidence, p component, and light polarized perpendicular to it, s component. At a particular angle, called the Brewster angle (also known as the polarization angle), no p

© 2015 by Taylor & Francis Group, LLC

polarized light is reflected from the surface; thus, all reflected light should be s polarized, with an electric field perpendicular to the plane of incidence. The Brewster law defines the relationship for light waves stating that the maximum polarization of a ray of light can be achieved by letting the ray fall on a surface of a transparent medium in a fashion that the refracted ray makes an angle of 90° with the reflected ray.

If a polarizer is used to analyze a polarized light, it is referred to as analyzer. It is an element whose intensity transmission is proportional to the content of a specific polarization state in the incident beam. The combination of a half-wave plate and the analyzer modifies the state of polarization of the incident beam of light appropriately. In general, the role of an analyzer can be performed by a calcite block introduced in the light path immediately after the stationary half-wave plate[84] and before a detecting device, such as a photo-multiplier tube ([162] and references therein) in polarimeters. If such an analyzer introduces equal amounts of amplitude change, it is known as a neutral density filter. A polarizer is characterized by the relations

$$\begin{aligned} E'_x &= p_x E_x \\ E'_y &= p_y E_y \qquad\qquad 0 \le p_{x,y} \le 1 \end{aligned} \tag{8.141}$$

where E'_x and E'_y are the components of the emerging beam, E_x and E_y are the components of the incident beam, and p_x and p_y the amplitude attenuation coefficients along orthogonal transmission axes, x and y, respectively.

For a perfect transmission, where there is no attenuation, along an orthogonal axis, $p_{x,y} = 1$, while for a complete attenuation, $p_{x,y} = 0$. If one of the axes has a zero absorption coefficient so that there is no transmission along this axis, the polarizer is said to have a single transmission axis. We have

$$\begin{aligned} S'_0 &= E'_x E'^*_x + E'_y E'^*_y \\ S'_1 &= E'_x E'^*_x - E'_y E'^*_y \\ S'_2 &= E'_x E'^*_y + E'_y E'^*_x \\ S'_3 &= i\left(E'_x E'^*_y - E'_y E'^*_x\right) \end{aligned} \tag{8.142}$$

in which E^* represents the complex conjugate of E.

Figure 8.11 depicts an input wave passing through an optical element, say, a polaroid sheet[85], with an electric field lying in the plane of the sheet. In general, a polaroid sheet is used as a linear polarizer and also as a retarder. The most common types of sheet polarizer are made by processes involving unidirectional stretching of large plastic sheets such as a polymer. These polarizers are thin and do not cause appreciable lateral shift of the image; therefore, it is possible to construct a simple astronomical polarimeter using these as analyzer. These polarizers are efficient in a restricted wavelength region, but have poor transmittance. If the field is decomposed with components parallel to and perpendicular to the stretched direction y, the field component in the case of the former suffers severe attenuation, while in the case of the

© 2015 by Taylor & Francis Group, LLC

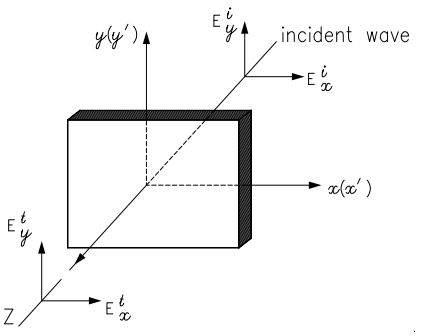


FIGURE 8.11: Schematic representation for optical element aligned with axes (x, y).

latter it suffers little attenuation. Consequently, the emerging wave is linearly polarized with the electric field perpendicular to the stretched direction.

8.4.3.2 Retarder

Another polarizing element of importance is the retarder, also called phase-shifter or compensator or wave plate. It introduces a differential phase-shift, $\delta = \delta_1 - \delta_2$, between the exiting beams for two orthogonal incident polarization states (the eigen-polarizations of the element). This can be thought of as being accomplished by causing a phase shift of $+\delta/2$ along the fast-axis (x axis) and a phase shift of $-\delta/2$ along the slow-axis (y axis). The fast-axis is the direction in which light travels faster, while the orthogonal direction is known as the slow-axis. For a positive uniaxial crystal the ellipsoid touches the spheroid along the former axis and for a negative uniaxial crystal along the latter axis. The retarder is, in the simplest case, a plane parallel plate of uniaxial crystal cut parallel to the optical axis. It divides the incident field into two components, for example, E_x and E_y, and retards the phase of one of these components relative to each other. When the two components of the wave are reunited to form a transmitted wave, the state of polarization is changed. In an ideal retarder, the phase of component E_y is retarded relative to the phase of the component E_x. In this case, the components of the emerging beam are related to the incident beam by

$$\begin{aligned} E'_x(z,t) &= E_x(z,t)e^{i\delta/2} \\ E'_y(z,t) &= E_y(z,t)e^{-i\delta/2} \end{aligned} \tag{8.143}$$

For an ideal phase-shifter (retarder), there is no loss in intensity, i.e., $S'_0 = S_0$. The most common types of retarders are

1. Quarter-wave plate: For a quarter-wave retarder, $\delta = 90°$, that is, the

© 2015 by Taylor & Francis Group, LLC

phase of one component of the light is delayed with respect to the orthogonal component by one quarter-wave. It is used to convert light between circular and linear polarization forms. Such a retarder changes linearly polarized light to circularly polarized light, when the angle between the input polarization and the retarder fast-axis is 45°. It can also be combined with a linear polarizer to form an optical isolator which is used to eliminate undesired reflections. A common application prevents unwanted reflected light from re-entering a laser cavity.

2. Half-wave plate: For a half-wave retarder, $\delta = 180°$, that is, the phase of one component of the light is delayed with respect to the orthogonal component by one half-wave. Such a retarder, also known as polarization rotator, flips the polarization direction of incoming light about the retarder fast-axis. When the angle between the retarder fast-axis and the input plane of polarization is 45°, horizontal polarized light is converted to vertical.

A retarder is naturally dependent on wavelength although there are achromatic retarders that are slowly dependent on wavelength.

8.4.3.3 Rotator

A rotator rotates the orthogonal components $E_x(z,t)$ and $E_y(z,t)$ by an angle ϕ; it changes the orientation angle of a polarization ellipse. Let us consider that an angle ϕ describes the rotation of E_x to E'_x and of E_y to E'_y. In Figure 8.12, the optical element is rotated by an angle ϕ about the z axis, so that $0 \leq \phi < \pi$. At the output of the rotator, the orthogonal components of the electric field, we may write

$$E'_x = E_x \cos\phi + E_y \sin\phi \tag{8.144}$$

$$E'_y = -E_x \sin\phi + E_y \cos\phi \tag{8.145}$$

Rotators are employed to change the orientation angle of the polarization ellipse (see equation 8.125). After passing through a rotator,

$$\begin{aligned} \tan 2\beta' &= \frac{-S_1 \sin 2\phi + S_2 \cos 2\phi}{S_1 \cos 2\phi + S_2 \sin 2\phi} = \frac{-\sin 2\phi + \tan 2\beta \cos 2\phi}{\cos 2\phi + \tan 2\beta \sin 2\phi} \\ &= \frac{\tan 2\beta - \tan 2\phi}{1 + \tan 2\beta \tan 2\phi} = \tan(2\beta - 2\phi) \end{aligned} \tag{8.146}$$

thus, $2\beta' = 2\beta - 2\phi$ or $\beta' = \beta - \phi$.

Equation (8.146) shows that rotators relate the polarization ellipse of the incident beam. The ellipticity remains unchanged. This expression is for clockwise rotation. If the rotation is counter clockwise, $\beta' = \beta + \phi$.

© 2015 by Taylor & Francis Group, LLC

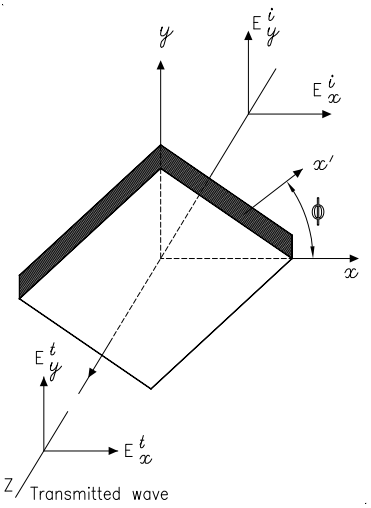


FIGURE 8.12: Rotation of optical field components by an optical element through angle ϕ, about γ axis; the optical element is rotated by the angle ϕ, about the z axis, so that $0 \leq \phi < \pi$.

8.4.4 Stokes Intensity Formula

The components E'_x and E'_y emerging from a retarder are described by equation (8.143). The component of E'_x along the transmission axis is $E'_x \cos\phi$, while the component of E'_y is $E'_y \sin\phi$. The field transmitted along this axis is the sum of these components; therefore, the total field E emerging from the polarizer is expressed as[124]

$$\begin{aligned} E &= E'_x \cos\phi + E'_y \sin\phi \\ &= E_x e^{i\delta/2} \cos\phi + E_y e^{-i\delta/2} \sin\phi \end{aligned} \tag{8.147}$$

As discussed in Section 1.5, the intensity I is given by equation (1.102), and so by taking the complex conjugate of equation (8.147), the intensity of the emerging beam is derived as

$$\begin{aligned} I(\phi,\delta) &= E_x E_x^* \cos^2\phi + E_y E_y^* \sin^2\phi \\ &\quad + E_x^* E_y e^{-i\delta} \sin\phi\cos\phi + E_x E_y^* e^{i\delta} \sin\phi\cos\phi \\ &= \frac{1}{2}\left[(E_x E_x^* + E_y E_y^*) + (E_x E_x^* - E_y E_y^*)\cos 2\phi \right. \\ &\quad \left. + (E_x E_y^* + E_y E_x^*)\cos\delta \sin 2\phi + i(E_x E_y^* - E_y E_x^*)\sin\delta \sin 2\phi\right] \end{aligned} \tag{8.148}$$

© 2015 by Taylor & Francis Group, LLC

in which $*$ represents the complex conjugate, ϕ the rotation angle of the transmission axis of the polarizer, δ the phase-shift of the retarder, and

$$\cos^2\phi = \frac{1+\cos 2\phi}{2}, \qquad \sin^2\phi = \frac{1-\cos 2\phi}{2}, \qquad \sin\phi\cos\phi = \frac{\sin 2\phi}{2}$$

On using equation (8.118), the Stokes intensity formula is recast as

$$I(\phi,\delta) = \frac{1}{2}\left[S_0 + S_1\cos 2\phi + S_2\cos\delta\sin 2\phi + S_3\sin\delta\sin 2\phi\right] \tag{8.149}$$

In order to describe unpolarized light using equation (8.149), Stokes observed that unpolarized light had a property, for instance, its intensity, that is unaffected by (i) rotation of a linear polarizer or (ii) the presence of a retarder. For such a light, the intensity $I(\phi,\delta)$ can be independent of ϕ,δ because this equation should satisfy $(\phi,\delta) = S_0/2)$ and $S_1 = S_2 = S_3 = 0$. The Stokes parameters, S_0, S_1, and S_2, are measured by removing the retarder ($\delta = 0°$) and rotating the transmission axis of the polarizer to the angles, $\phi = 0°$, $+45°$, and, $+90°$, respectively[124], while S_3 is measured by reinserting a quarter-wave plate ($\delta = 90°$) into the optical path and setting the transmission axis of the polarizer to $\phi = 45°$. The intensities are found to be

$$\begin{aligned} I(0°,0°) &= \frac{1}{2}\left[S_0 + S_1\right] \\ I(45°,0°) &= \frac{1}{2}\left[S_0 + S_2\right] \\ I(90°,0°) &= \frac{1}{2}\left[S_0 - S_1\right] \\ I(45°,90°) &= \frac{1}{2}\left[S_0 + S_3\right] \end{aligned} \tag{8.150}$$

In order to measure Stokes parameters, it is prudent to measure the intensity at four angles, which can be obtained by solving equation (8.150):

$$\begin{aligned} S_0 &= I(0°,0°) + I(90°,0°) \\ S_1 &= I(0°,0°) - I(90°,0°) \\ S_2 &= 2I(45°,0°) - I(0°,0°) - I(90°,0°) \\ S_3 &= 2I(45°,90°) - I(0°,0°) - I(90°,0°) \end{aligned} \tag{8.151}$$

8.4.5 Jones Matrix

In order to facilitate the treatment of complicated problems at the amplitude level, R. Clerk Jones, in the early 1940s, developed a matrix calculus, which is referred to as the Jones matrix calculus[86]. It can be defined for polarizing optical components allow the nature of the polarization to propagate easily. Let us assume that the components of a beam emerging from a

© 2015 by Taylor & Francis Group, LLC

polarizing element are linearly related to the components of the incident light that propagates as a polarized monochromatic plane wave. These elements change the state of polarization of the incident beam of light appropriately, but do not affect the direction of propagation. This relation is expressed as

$$E'_x = J_{xx}E_x + J_{xy}E_y \tag{8.152}$$
$$E'_y = J_{yx}E_x + J_{yy}E_y \tag{8.153}$$

where the quantities, J_{ij} are the transforming factors (elements), $i, j = x, y$; E'_x and E'_y the components of the emerging beam; and E_x and E_y the components of the incident beam.

For a linear optical element, the relation between the input and output Jones vectors, $\mathbf{E}$ and $\mathbf{E}'$, is expressible in the form

$$\begin{aligned} \mathbf{E}' &= \mathbf{J}\cdot\mathbf{E} = \begin{bmatrix} E'_x \\ E'_y \end{bmatrix} \\ &= \begin{bmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix} \end{aligned} \tag{8.154}$$

in which the 2×2 matrix, called the Jones matrix (or coherent) matrix, is

$$\mathbf{J} = \begin{bmatrix} \langle E_x E_x^* \rangle & \langle E_x E_y^* \rangle \\ \langle E_y E_x^* \rangle & \langle E_y E_y^* \rangle \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{bmatrix} \tag{8.155}$$

which couples the incident wave of the system to the transmitted wave of that system.

8.4.5.1 Jones Matrix for a Polarizer

In terms of the Jones vector, equation (8.154) translates into

$$\begin{bmatrix} E'_x \\ E'_y \end{bmatrix} = \begin{bmatrix} p_x & 0 \\ 0 & p_y \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix} \tag{8.156}$$

where the Jones matrix for a polarizer is

$$\mathbf{J}_p = \begin{bmatrix} p_x & 0 \\ 0 & p_y \end{bmatrix} \qquad 0 \leq p_{x,y} \leq 1 \tag{8.157}$$

An ideal linear polarizer is one which has transmission along one axis and no transmission along the orthogonal axis. We discuss below two related Jones matrices:

1. Ideal linear horizontal polarizer (ph):

$$\mathbf{J}_{ph} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{8.158}$$

© 2015 by Taylor & Francis Group, LLC

2. Ideal linear vertical polarizer (pv):

$$\mathbf{J}_{pv} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{8.159}$$

It is essential to know the Jones matrix for a linear polarizer rotated through an angle, ϕ, by employing rotation transformation. The action of rotating such a polarizer through this angle with respect to the one direction results in

$$\mathbf{J}' = \mathbf{J}(-\phi) \cdot \mathbf{J} \cdot \mathbf{J}(\phi) \tag{8.160}$$

Here $\mathbf{J}$ is given by equation (8.155). The rotation matrix, $\mathbf{J}(-\phi)$, in the other direction is

$$\mathbf{J}(-\phi) = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \tag{8.161}$$

Here $\mathbf{R}_l^{-1} = \mathbf{R}_r$, where $\mathbf{R}_l$ is the matrix introduced for rotating the linear polarizer through the angle ϕ with respect to the other direction.

For the cascaded elements, the transmission matrix is the product of the transmission matrices for the individual elements

$$\mathbf{J}' = \mathbf{J}_1 \cdot \mathbf{J}_2 \cdots \mathbf{J}_{m-1} \cdot \mathbf{J}_m \tag{8.162}$$

The form of the Jones matrix for a linear polarizer is given by

$$\mathbf{J}' = \begin{bmatrix} J_{xx} & J_{xy} \\ 0 & 0 \end{bmatrix} \tag{8.163}$$

With additional constraints that an input wave linearly polarized along the transmission (x axis) is transmitted unattenuated and that an input wave linearly polarized along the y axis normal to the transmission axis is completely absorbed, the co-efficients, J_{xx} and J_{xy}, are determined by

$$\mathbf{J} = e^{-i\xi} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{8.164}$$

in which ξ is the phase shift on transmission.

The Jones matrix of an ideal linear polarizer after removing common phase factor $e^{-i\xi}$ is normalized to equation (8.158). When the transmission axis (x axis) of the linear polarizer represented by equation (8.163) is at the angle ϕ, as in Figure 8.12, the Jones matrix from equation (8.160) is

$$\mathbf{J}' = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} p_x & 0 \\ 0 & p_y \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \tag{8.165}$$

Also, the Jones matrix for a rotated polarizer is

$$\mathbf{J}'(\phi) = \begin{bmatrix} p_x\cos^2\phi + p_y\sin^2\phi & (p_x - p_y)\sin\phi\cos\phi \\ (p_x - p_y)\sin\phi\cos\phi & p_x\sin^2\phi + p_y\cos^2\phi \end{bmatrix} \tag{8.166}$$

© 2015 by Taylor & Francis Group, LLC

For an ideal linear horizontal polarizer, we may set $p_x = 1$ and $p_y = 0$ in equation (8.166), so that the Jones matrix for a rotated linear horizontal polarizer is

$$\mathbf{J}_p(\phi) = \begin{bmatrix} \cos^2\phi & \cos\phi\sin\phi \\ \cos\phi\sin\phi & \sin^2\phi \end{bmatrix} \tag{8.167}$$

The Jones matrix having a linear polarizer rotated though $+45^\circ$ is envisaged from equation (8.167) as

$$\mathbf{J}_p(45^\circ) = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{8.168}$$

while in the case of the nonideal linear polarizer, the Jones matrix for a polarizer at 45° is seen from equation (8.166) as

$$\mathbf{J}_p(45^\circ) = \frac{1}{2}\begin{bmatrix} p_x + p_y & p_x - p_y \\ p_x - p_y & p_x + p_y \end{bmatrix} \tag{8.169}$$

For a neutral density filter (ND; see Section 8.4.3.1), $p_x = p_y = p$; hence,

$$\mathbf{J}_{\mathrm{ND}}(\phi) = p\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{8.170}$$

The Jones vector for the transmitted wave (equation 8.144) is deduced as

$$\mathbf{E}' = \mathbf{J}\cdot\mathbf{E} = E'_x e^{-i\xi}\cos\phi\begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \tag{8.171}$$

Thus, the Malus law for the linear polarizer is determined by the ratio of the time average powers per unit area for the transmitted and incident wave. If an unpolarized light passes through a polarizer, the intensity of the transmitted light, I, would be $1/2$ of what it is coming in. According to Malus, when linearly polarized light passes through a polarizer, the transmitted light intensity is directly proportional to the square of the cosine of the angle between the transmission axes of the analyzer and the polarizer, i.e., $S'_0 \propto \cos^2\phi$, in which ϕ is the angle between the plane of vibration and the polarizing direction of the sheet, $S_0(\propto E_0^2)$ the initial intensity, and E_0 the amplitude of the electric vector transmitted by the polarizer. Let an incident plane-polarized wave, $\mathbf{E} = \mathbf{E}_0\sin(\omega t + \psi)$, make an angle with the transmission axis. The electric field vector, $\mathbf{E}_0$, can be resolved into two rectangular components, $\mathbf{E}_0\cos\phi$ and $\mathbf{E}_0\sin\phi$. The analyzer transmits component $E_0\cos\phi$ which is parallel to its transmission axis. However, component $E_0\sin\phi$ is absorbed by the analyzer, and therefore,

$$\frac{I}{I_0} = \frac{(\mathbf{E}_0\cos\phi)\cdot(\mathbf{E}_0\cos\phi)}{\mathbf{E}_0\cdot\mathbf{E}_0} = \cos^2\phi \tag{8.172}$$

When $\phi = 0^\circ$ (or 180°), $S'_0 = S_0$ which means the intensity of light transmitted by the analyzer is maximum when the transmission axes of the analyzer

© 2015 by Taylor & Francis Group, LLC

and the polarizer are parallel. If $\phi = 90°$, $S_0' = 0$, i.e., the intensity of light transmitted by the analyzer is minimum when the transmission axes of the analyzer and polarizer are perpendicular to each other.

Let us express equation (8.156) in the Jones transformation as

$$\begin{aligned} \mathbf{J}' &= \begin{bmatrix} E_x' \\ E_y' \end{bmatrix} \\ &= \begin{bmatrix} e^{i\delta/2} & 0 \\ 0 & e^{-i\delta/2} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix} \end{aligned} \tag{8.173}$$

The Jones matrix for a retarder, $\mathbf{J}_{\mathrm{R}}$, is

$$\mathbf{J}_{\mathrm{R}} = \begin{bmatrix} e^{i\delta/2} & 0 \\ 0 & e^{-i\delta/2} \end{bmatrix} \tag{8.174}$$

in which δ is the total phase shift between the field components.

8.4.5.2 Jones Matrix for a Retarder

As discussed in Section 8.4.3.2, for the quarter-wave retarder and the half-wave retarder, the phase-shifts are $\delta = \pi/2$ and $\delta = \pi$, respectively; hence, the Jones matrices for these elements are expressed as

$$\begin{aligned} \mathbf{J}_{\mathrm{R}}\left(\frac{1}{4}\right) &= \begin{bmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{bmatrix} = e^{i\pi/4} \begin{bmatrix} 1 & 0 \\ 0 & e^{-i\pi/2} \end{bmatrix} = e^{i\pi/4} \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix} \\ \mathbf{J}_{\mathrm{R}}\left(\frac{1}{2}\right) &= \begin{bmatrix} e^{i\pi/2} & 0 \\ 0 & e^{-i\pi/2} \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} = i \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \end{aligned} \tag{8.175}$$

The Jones matrix for a rotated retarder is found from equations (8.160 and 8.174) as

$$\mathbf{J}_{\mathrm{R}}(\delta, \phi) = \begin{bmatrix} e^{i\delta/2}\cos^2\phi + e^{-i\delta/2}\sin^2\phi & \left(e^{i\delta/2} - e^{-i\delta/2}\right)\sin\phi\cos\phi \\ \left(e^{i\delta/2} - e^{-i\delta/2}\right)\sin\phi\cos\phi & e^{i\delta/2}\sin^2\phi + e^{-i\delta/2}\cos^2\phi \end{bmatrix} \tag{8.176}$$

When the fast-axis (x' axis) of the linear retarder is at the angle ϕ as depicted in Figure 8.12, the normalized Jones matrix in the case of a quarter-wave plate translates into

$$\mathbf{J}_{\mathrm{R}} = \begin{bmatrix} \cos(\pi/4) + i\cos(2\phi)\sin(\pi/4) & i\sin(2\phi)\sin(\pi/4) \\ \mp i\sin(2\phi)\sin(\pi/4) & \cos(\pi/4) - i\cos(2\phi)\sin(\pi/4) \end{bmatrix} \tag{8.177}$$

An ideal circular polarizer produces either a left-handed (LCP) or right-handed circularly polarized (RCP) transmitted field from an incident field of arbitrary polarization. A circular polarizer is generally made from a linear polarizer placed between two quarter-wave plates as in Figure 8.13, where the angles of the linear polarizer, $\phi = 45°$, and $\phi = -45°$ are maintained for the respective RCP and LCP. The incident light (1) with arbitrary polarization

© 2015 by Taylor & Francis Group, LLC

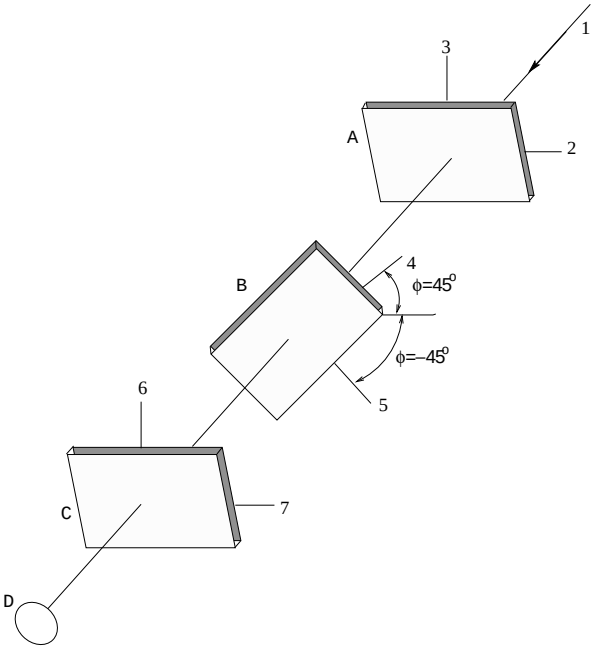


FIGURE 8.13: Three-element, circular polarizer formed from two quarter-wave plates `A`, `C` and a linear polarizer, `B`. Angle of linear polarizer; $\phi = 45°$ for LCP and $\phi = -45°$ for RCP.

is passing through a quarter-wave plate (`A`). The fast and slow axes are represented by numbers 2 and 3, respectively, while the numbers 4 and 5 are the transmission axes for the LCP which is also a horizontal linear analyzer, and right-handed circular polarizer (RCP), respectively. The wave emerges out of a linear polarizer (`B`) and passes through another quarter-wave plate (`C`) as a circularly polarized wave (`D`), in which the numbers 6 and 7 are fast and slow axes, respectively. By multiplying Jones matrices for these elements, we find

$$\mathbf{J}'_{\mathrm{R}} = \begin{bmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & \pm 1 \\ \pm 1 & 1 \end{bmatrix} \begin{bmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & \pm i \\ \mp i & 1 \end{bmatrix} \tag{8.178}$$

where the sign $\pm$ refers to right-handed and left-handed circular polarization.

This matrix is the normalized Jones matrix for right-handed (left-handed) circular polarizer. The changes in the state of polarization for a wave passing through the circular polarizer can be witnessed by the Poincaré system[108].

8.4.5.3 Jones Matrix for a Rotator

Another matrix of interest is the Jones matrix for a rotator, for which the defining equations are given by equation (8.144), in which ϕ is the angle of

© 2015 by Taylor & Francis Group, LLC

rotation; hence,

$$\mathbf{J}' = \begin{bmatrix} E'_x \\ E'_y \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix} \tag{8.179}$$

The Jones matrix for a rotator is represented by a unitary rotation matrix defined for the linear polarization-state basis set as

$$\mathbf{J}_{rot} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \tag{8.180}$$

8.4.6 Mueller Matrix

The interaction of polarized light with elements can change its state of polarization, which can be represented in a matrix form. This matrix representation leads to a very powerful mathematical tool, named after Hans Mueller (1900−1965), called the Mueller matrix[87]. It is a generalization of the Jones matrix (see Section 8.4.5). However, the description of optical systems in terms of the former is applicable to more a general situation than is the description in terms of the latter. The Mueller matrix is a 4×4 matrix which can be used to reproduce the effect of a given optical element when applied to a Stokes vector, S. The effect of an optical device that transforms the aberrations into the light intensity variations, on the polarization of light, is characterized by such a matrix. Light which is unpolarized or partially polarized must be treated using Mueller calculus, while fully polarized light may be treated with Jones calculus since the latter works with amplitude rather than intensity of light.

Let an incident beam interacting with a polarizing element be characterized by its Stokes parameters, S_0, S_1, S_2, and S_3. The input polarized beam interacts with the polarizing medium, and hence, the output beam is characterized by a new set of Stokes parameters, S'_0, S'_1, S'_2, and S'_3:

$$\begin{aligned} S'_0 &= m_{00}S_0 + m_{01}S_1 + m_{02}S_2 + m_{03}S_3 \\ S'_1 &= m_{10}S_1 + m_{11}S_1 + m_{12}S_2 + m_{13}S_3 \\ S'_2 &= m_{20}S_2 + m_{21}S_1 + m_{22}S_2 + m_{23}S_3 \\ S'_3 &= m_{30}S_3 + m_{31}S_1 + m_{32}S_2 + m_{33}S_3 \end{aligned} \tag{8.181}$$

In terms of the Stokes column matrix, we write

$$\begin{aligned} S' &= \begin{bmatrix} S'_0 \\ S'_1 \\ S'_2 \\ S'_3 \end{bmatrix} = \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \\ m_{30} & m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \\ &= M \cdot S \end{aligned} \tag{8.182}$$

© 2015 by Taylor & Francis Group, LLC

in which

$$M = \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \\ m_{30} & m_{31} & m_{32} & m_{33} \end{bmatrix} \tag{8.183}$$

is known as the Mueller matrix of the optical element, which depends on the frequency.

The Mueller matrix M', associated with a beam path through a sequence of polarization elements, is

$$M' = M_1 \cdot M_2 \cdots M_{m-1} \cdot M_m \tag{8.184}$$

The Mueller matrix exhibits an angular dependence in the case of light scattering.

8.4.6.1 Mueller Matrix of a Polarizer

Substituting equation (8.141) into (8.142) and using equation (8.118), we find

$$\begin{bmatrix} S_0' \\ S_1' \\ S_2' \\ S_3' \end{bmatrix} = \frac{1}{2} \begin{bmatrix} p_x^2 + p_y^2 & p_x^2 - p_y^2 & 0 & 0 \\ p_x^2 - p_y^2 & p_x^2 + p_y^2 & 0 & 0 \\ 0 & 0 & 2p_x p_y & 0 \\ 0 & 0 & 0 & 2p_x p_y \end{bmatrix} \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \tag{8.185}$$

Thus, the Mueller matrix for a polarizer with amplitude attenuation coefficients, p_x and p_y, can be expressed as[124]

$$M = \frac{1}{2} \begin{bmatrix} p_x^2 + p_y^2 & p_x^2 - p_y^2 & 0 & 0 \\ p_x^2 - p_y^2 & p_x^2 + p_y^2 & 0 & 0 \\ 0 & 0 & 2p_x p_y & 0 \\ 0 & 0 & 0 & 2p_x p_y \end{bmatrix} \qquad 0 \le p_x, p_y \le 1 \tag{8.186}$$

The Mueller matrices in the following cases are enumerated in brief:

1. Neutral density filter: For a neutral density filter, $p_x = p_y = p$; therefore,

$$M_{\mathrm{ND}} = p^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8.187}$$

   So, for an incident beam of intensity, I, falling on a neutral density filter, the output intensity is $I' = p^2 I$.

2. Horizontal polarizer: For horizontal polarizer, $p_y = 0$; hence,

$$M_{ph} = \frac{p_x^2}{2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{8.188}$$

© 2015 by Taylor & Francis Group, LLC

For ideal horizontal polarizer, $p_x = 1$ and $p_y = 0$; thus,

$$M_{ph} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{8.189}$$

3. Vertical polarizer: For a vertical polarizer, $p_x = 0$, so

$$M_{pv} = \frac{p_y^2}{2}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{8.190}$$

In the case of an ideal vertical polarizer, $p_x = 0$ and $p_y = 1$; hence,

$$M_{pv} = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{8.191}$$

The form of the Mueller matrix of an ideal linear polarizer oriented at angle ϕ is expressed as[115]

$$M_P(\phi) = \frac{1}{2}\begin{bmatrix} 1 & \cos 2\phi & \sin 2\phi & 0 \\ \cos 2\phi & \cos^2 2\phi & \frac{1}{2}\sin 4\phi & 0 \\ \sin 2\phi & \frac{1}{2}\sin 4\phi & \sin^2 2\phi & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{8.192}$$

8.4.6.2 Mueller Matrix of a Retarder

The Mueller matrix for an ideal homogeneous linear compensator can be derived from the following expressions that arise from the definition of Stokes parameters (equation 8.142):

$$\begin{aligned} S_0' &= |E_x'|^2 + |E_y'|^2 = |E_x|^2 + |E_y|^2 = S_0 \\ S_1' &= |E_x'|^2 - |E_y'|^2 = |E_x|^2 - |E_y|^2 = S_1 \\ S_2' &= E_x' E_2'^* + E_y' E_1'^* = S_2 \cos\delta + S_3 \sin\delta \\ S_3' &= i(E_x' E_y'^* - E_y' E_x'^*) = -S_2 \sin\delta + S_3 \cos\delta \end{aligned} \tag{8.193}$$

Equation (8.193) is expressed in matrix form as

$$\begin{bmatrix} S_0' \\ S_1' \\ S_2' \\ S_3' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\delta & \sin\delta \\ 0 & 0 & -\sin\delta & \cos\delta \end{bmatrix} \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \tag{8.194}$$

© 2015 by Taylor & Francis Group, LLC

There is no loss in intensity, $S'_0 = S_0$; thus, Mueller matrix for a phase-shifter with a phase-shift δ is

$$M_{Re} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\delta & \sin\delta \\ 0 & 0 & -\sin\delta & \cos\delta \end{bmatrix} \tag{8.195}$$

For any beam of light, the quantities $S_0, (S_1^2 + S_2^2)$ and S_3 are invariant under a rotation of the reference coordinate system.

8.4.6.3 Mueller Matrix of a Rotator

In the case of a rotator, we write the following expression as

$$\begin{aligned} S'_0 &= \mid E'_x \mid^2 + \mid E'_y \mid^2 = \mid E_x \mid^2 + \mid E_y \mid^2 = S_0 \\ S'_1 &= \mid E'_x \mid^2 - \mid E'_y \mid^2 = \mid E_x \mid^2 - \mid E_y \mid^2 = S_1 \cos 2\phi + S_2 \sin 2\phi \\ S'_2 &= E'_x E'^*_y + E'_y E'^*_x = -S_2 \sin 2\phi + S_3 \cos 2\phi \\ S'_3 &= i(E'_x E'^*_y - E'_y E'^*_x) = S_3 \end{aligned} \tag{8.196}$$

Thus, the Mueller matrix of a rotator is expressed as

$$M_R(\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\phi & \sin 2\phi & 0 \\ 0 & -\sin 2\phi & \cos 2\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8.197}$$

When a polarization element with the Mueller matrix, M, is rotated about the beam of light by an angle, ϕ, such that the angle of incidence is unchanged (for instance, for a normal-incidence beam, rotating the element about the normal), the resulting Mueller matrix $M(\phi)$ is

$$\begin{aligned} M(\phi) &= M_R(-\phi) \cdot M \cdot M_R(\phi) \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\phi & -\sin 2\phi & 0 \\ 0 & \sin 2\phi & \cos 2\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \\ m_{30} & m_{31} & m_{32} & m_{33} \end{bmatrix} \\ &\quad \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\phi & \sin 2\phi & 0 \\ 0 & -\sin 2\phi & \cos 2\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{8.198}$$

A rotated Mueller matrix expressed by equation (8.198) appears in the treatment of polarized light, particularly in deriving Mueller matrices for a rotated polarizer and a rotated retarder[124]. From equation (8.192), we find that Stokes parameters S'_0, S'_1, S'_2, and S'_3 of light transmitted through a

© 2015 by Taylor & Francis Group, LLC

perfect analyzer with the principal plane containing the optical axis and the incident beam, at position angle, ϕ, are related to Stokes parameters S_0, S_1, S_2, and S_3 of the incident light by the Mueller matrix of transformation: equation[163]

$$\begin{bmatrix} S_0' \\ S_1' \\ S_2' \\ S_3' \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & \cos 2\phi & \sin 2\phi & 0 \\ \cos 2\phi & \cos^2 2\phi & \frac{1}{2}\sin 4\phi & 0 \\ \sin 2\phi & \frac{1}{2}\sin 4\phi & \sin^2 2\phi & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \tag{8.199}$$

Equation (8.199) reveals that Stokes parameters S_0, S_1, S_2, and S_3 of the incident light can be determined by measuring the intensities of light transmitted by an analyzer with its optical axis at several position angles ϕ. If the analyzer is rotated at a frequency ν, the transmitted light modulates at a frequency 2ν. Multiplying the expression for Stokes vector of an incident light (equation 8.113) by equation (8.197) yields the emerging Stokes vector, S', that is,

$$S' = \begin{bmatrix} S_0 \\ S_1 \cos 2\phi + S_2 \sin 2\phi \\ -S_1 \sin 2\phi + S_2 \cos 2\phi \\ S_3 \end{bmatrix} \tag{8.200}$$

The ellipticity angle, χ', is

$$\sin 2\chi' = \frac{S_3'}{S_0'} = \frac{S_3}{S_0} = \sin 2\chi \tag{8.201}$$

For a true rotator, the orientation angle β is derived as

$$\begin{aligned} \tan 2\beta' &= \frac{S_2'}{S_1'} = \frac{\sin 2\beta \cos 2\phi - \sin 2\phi \cos 2\beta}{\cos 2\beta \cos 2\phi + \sin 2\beta \sin 2\phi} \\ &= \frac{\sin(2\beta - 2\phi)}{\cos(2\beta - 2\phi)} \end{aligned} \tag{8.202}$$

whence $\beta' = \beta - \phi$.

8.4.7 Rotated Polarizing Elements

In order to derive the Mueller matrix for rotated polarizing components, as stated in Section 8.4.6.3, the axes of the polarizing component are rotated through an angle to the x' and y' axes. In terms of the Stokes vector of the incident beam S, equation (8.192) is $S' = M_R(\phi).S$, in which $M_R(\phi)$ represents the Mueller matrix for rotation (see equation 8.197). The S' beam interacts with a polarizing element characterized by its Mueller matrix M; the Stokes vector S'' of the emerging from the rotated polarizing component becomes $S'' = M.S' = M.M_R(\phi).S$. If we take the components of the emerging

© 2015 by Taylor & Francis Group, LLC

beam along the original x and y axes, which is counterclockwise rotation of S'' through $-\phi$ and back to the initial x, y axes (see Figure 8.13), we get

$$S''' = [M_R(-\phi).M.M_R(\phi)].S = M(\phi).S \tag{8.203}$$

with $M(\phi) = M_R(-\phi).M.M_R(\phi)$.

The rotated Mueller matrix given by equation (8.203) appears in the treatment of polarized light, generally in the case of Mueller matrices for a rotated polarizer and a rotated retarder. On carrying out the matrix multiplication given by this expression, the Mueller matrix for the rotated retarder (equation 8.195) is given by

$M_{Re}(\phi, \delta) =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2 2\phi + \cos\delta \sin^2 2\phi & (1-\cos\delta)\sin 2\phi \cos 2\phi & -\sin\delta \sin 2\phi \\ 0 & (1-\cos\delta)\sin 2\phi \cos 2\phi & \sin^2 2\phi + \cos\delta \cos^2 2\phi & \sin\delta \cos 2\phi \\ 0 & \sin\delta \sin 2\phi & -\sin\delta \cos 2\phi & \cos\delta \end{bmatrix} \tag{8.204}$$

For the Mueller matrix for the rotated retarder, M_{Re} in the case of $\phi = 0°$, equation (8.204) reduces to equation (8.195). In what follows, we enumerate two important cases.

8.4.7.1 Quarter-Wave Retarder

In the case of a quarter-wave retarder, the phase-shift is $\delta = \pi/2$ (see Section 8.4.3.2); thus, the Mueller matrix becomes

$$M_{Re}\left(\frac{1}{4}\right) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2 2\phi & \sin 2\phi \cos 2\phi & -\sin 2\phi \\ 0 & \sin 2\phi \cos 2\phi & \sin^2 2\phi & \cos 2\phi \\ 0 & \sin 2\phi & -\cos 2\phi & 0 \end{bmatrix} \tag{8.205}$$

For a linear horizontal polarized light, the Stokes vector S is

$$S = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \tag{8.206}$$

and

$$S' = \begin{bmatrix} 1 \\ \cos^2\phi \\ \sin 2\phi \cos 2\phi \\ \sin 2\phi \end{bmatrix} \tag{8.207}$$

We see that $\tan 2\beta' = \tan 2\phi$ and $\sin 2\chi' = \sin 2\phi$. The rotated quarter-wave plate generates a polarization ellipse of desired orientation and ellipticity commencing with an incident horizontally polarized beam.

© 2015 by Taylor & Francis Group, LLC

8.4.7.2 Half-Wave Retarder

In the case of a half-wave retarder, the phase shift is $\delta = \pi$ (see Section 8.4.3.2); hence, the Mueller matrix turns out to be

$$M_{Re}\left(\frac{1}{2}\right) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 4\phi & \sin 4\phi & 0 \\ 0 & \sin 4\phi & -\cos 4\phi & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{8.208}$$

The ellipticity angle χ' given by equation (8.201) is not changed under true rotation. But for a rotated half-wave retarder, the Stokes vector S' is obtained by multiplying equation (8.113) by equation (8.208):

$$S' = \begin{bmatrix} S_0 \\ S_1 \cos 4\phi + S_2 \sin 4\phi \\ S_1 \sin 4\phi - S_2 \cos 4\phi \\ -S_3 \end{bmatrix} \tag{8.209}$$

Here, the ellipticity angle χ' is

$$\sin 2\chi' = \frac{S_3'}{S_0'} = \frac{-S_3}{S_0} = -\sin 2\chi \tag{8.210}$$

so $\chi' = \chi + 90°$, which means that the ellipticity angle χ of the incident beam is advanced 90° by means of a rotated half-wave retarder.

For rotated retarder, the orientation angle β is derived as

$$\begin{aligned} \tan 2\beta' &= \frac{\cos 2\beta \sin 4\phi - \sin 2\beta \cos 4\phi}{\cos 2\beta \cos 4\phi + \sin 2\beta \sin 4\phi} \\ &= \frac{\sin(4\phi - 2\beta)}{\cos(4\phi - 2\beta)} \end{aligned} \tag{8.211}$$

whence $\beta' = 2\phi - \beta$.

Rotating a half-wave retarder clockwise causes β' to rotate counterclockwise by an amount twice that of a rotator.

8.4.8 Elliptical and Circular Polarizer and Analyzer

The linear polarizers (LP; see Section 8.4.3.1) are simple and inexpensive; therefore, many circular and elliptical polarizers incorporate these devices to perform the polarizing and retarders to convert polarization states. For such compound devices, the distinction between a polarizer and an analyzer becomes significant. In what follows we describe some examples.

© 2015 by Taylor & Francis Group, LLC

8.4.8.1 Left-Handed Circular Polarizer

A left-handed circular polarizer (LCP) is, in general, constructed from a horizontal linear polarizer having 0° axis followed by a quarter-wave linear retarder (QWLR) with the fast-axis oriented at 135°, that is,

$$\begin{aligned} QWLR(135^\circ)LP(0^\circ)S &= \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 \end{bmatrix}\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} S_0 + S_1 \\ 0 \\ 0 \\ -S_0 - S_1 \end{bmatrix} \end{aligned} \tag{8.212}$$

Equation (8.212) states that the device transmits left-handed circularly polarized light, since the zeroth and third elements have equal magnitude and opposite sign, making it an LCP. However, the transmitted flux $(S_0 + S_1)/2$ is the flux of horizontal linearly polarized light in the incident beam, which makes it a horizontal linear analyzer.

8.4.8.2 Left-Handed Circular Analyzer

A left-handed circular analyzer is constructed from a QWLR having axis 45° followed by a linear polarizer with 0°; that is,

$$\begin{aligned} LP(0^\circ)QWLR(45^\circ)S &= \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} S_0 - S_3 \\ S_0 - S_3 \\ 0 \\ 0 \end{bmatrix} \end{aligned} \tag{8.213}$$

Equation (8.213) explains that the transmitted flux, $(S_0 - S_3)/2$, is the flux of left-handed circularly polarized light in the incident beam making this combination a left-handed circular analyzer. The final polarizer makes the device a horizontal linear polarizer. Thus, an analyzer for a state does not necessarily transmit the state; its transmitted flux is proportional to the amount of the analyzed state in the incident beam. Equations (8.212 and 8.213) are known as inhomogeneous polarization elements because the eigenpolarizations are not orthogonal, and the characteristics of the device are different for propagation in opposite directions.

© 2015 by Taylor & Francis Group, LLC

8.4.8.3 Left Circular Analyzer and Polarizer

This system is made up of a QWLR with 135°, a linear polarizer with 0° axis, followed by a QWLR having axis 45°; that is,

$$\begin{aligned} QWLR(135^\circ)LP(0^\circ)QWLR(45^\circ)S &= \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} S_0 - S_3 \\ 0 \\ 0 \\ -S_0 + S_3 \end{bmatrix} \end{aligned} \tag{8.214}$$

The device in equation (8.214) is both a left-handed circular polarizer and a left-handed circular analyzer having the same characteristics for propagation in opposite directions. Such a device is referred to as a homogeneous left-handed circular polarizer.

8.4.9 Polarimetry

Polarimetry is the science of measuring the polarization state of a light beam and the diattenuating, retarding, and depolarizing properties of materials. A polarimetric technique that uses the change in the state of polarization of light upon reflection for the characterization of surfaces, interfaces, and thin films is referred to as ellipsometry. Such a method can be used in remote sensing applications, such as, planetary science and weather radar. This can also be included in computational analysis of waves. For instance, radars often consider wave polarization in post-processing to enhance the characterization at the targets. In this case, polarimetry can be used to estimate the fine texture of a material, help to resolve the orientation of small structures in the target, and when the circularly polarized antennas are used, to resolve the number of bounces of the received signal. Another field of importance is the spectroscopic study of the polarization properties of materials, what is referred to as the spectropolarimetry. A generalization of conventional optical spectroscopy (see Section 1.2), it determines the diattenuating, retarding, and depolarizing properties of the sample. Characterization of these properties is accomplished by measuring the Mueller matrix of the sample as a function of wavelength.

8.4.9.1 Polarimeter

A polarimeter is the basic scientific equipment for the determination of the polarization state of a light beam or the polarization-altering properties of a sample. Polarimetry can be used to measure the polarization state of a light beam and the diattenuating, retarding, and depolarizing properties of

© 2015 by Taylor & Francis Group, LLC

materials; linear and circular birefringences, linear and circular dichroisms, and scattering can also be measured. The most sensitive polarimeters are based on interferometers, while the conventional polarimeters are based on arrangements of polarizing filters, wave-plates, or other devices.

Optically active samples, such as chiral molecules, often exhibit circular birefringence causing rotation of the polarization of plane polarized light as it passes through the sample. In an ordinary light that obeys the laws of refraction (Snell law), the vibration occurs in all planes perpendicular to the direction of propagation. When it is allowed to pass through a Nicol prism, its vibrations in all directions baring the direction of axis of the prism are cut-off. The light emerging from the prism is said to be plane polarized since its vibration is in one direction. If two such prisms are placed with their planes of polarization parallel to each other, the light rays emerging from the first prism, called polarizer, would enter the second prism, which is referred to as analyzer, as a result of which complete bright light would be observed. If the second prism is rotated by an angle of 90°, the light emerging from the first prism is blocked by the second prism and complete dark (no light region) is observed. Since the optical components, such as, dichroic filters and diffraction grating, change the state of polarization of the incident beam, the analyzer should be placed in front of all such components.

A simple polarimeter to measure this rotation consists of a long tube with flat glass ends, into which the sample is placed. At each end of the tube is a Nicol prism or any other polarizer. Light is shone through the tube and the prism at the other end, attached to an eyepiece, and rotated to measure the region of complete dark region. The angle of rotation is read from a scale. The same phenomenon is observed after an angle of 180°. The specific rotation at the sample may then be calculated. Temperature can affect the rotation of light, which should be accounted for in the calculations. The specific rotation is given by

$$[\alpha]_{\lambda}^{T} = \frac{100\alpha}{L\rho} \tag{8.215}$$

with T as the temperature, λ the wavelength of light, α the angle of rotation, L the length of polarimeter, and ρ the mass concentration of solution.

8.4.9.2 Astronomical Polarimeter

The light of stars is unpolarized, albeit environmental effects such as scattering by aligned dust particles or free electrons, or anisotropic scattering may polarize the light. Various absorption and emission processes in the presence of a magnetic field can produce polarized radiation, for instance, synchroton emission. Studies of the dependence of the polarization on wavelength and time help to construct models of the source that explain the behavior of the source.

In most cases, the degree of polarization is on the order of a few percent. This requires measurements of the signal-to-noise (S/N) ratio on the order

© 2015 by Taylor & Francis Group, LLC

of 1000 and hence, an effective measuring device. Also, it requires a careful calibration of instrumental polarization effects, which is done by observations of polarimetric standards.

The simplest form of an imaging polarimeter would be the insertion of a polaroid film into the light beam and the measurement of the same field with four different orientations of the polaroid. However, this method is obviously very sensitive to fluctuations in atmospheric transparency and, therefore, of limited use.

In an imaging polarimeter, a birefringent polarizer Wollaston prism is used as a polarizer that splits the image of an object into two perpendicularly polarized images and a rotatable half-wave plate to change the polarization angle. Rotating the half-wave plate a 22.5° step would provide the position angle 0 and 90° in the first exposure, 45 and 135° in the second, 90 and 180° in the third, and 135 and 225° in the fourth exposure. A strip mask is placed in the focal plane to the overlapping of the two images. If a quarter-wave plate is inserted in the beam, the circular polarized beam is converted into a linear polarized beam.

In an astronomical polarimeter[84], a beam displacement prism divides the incident light into two beams with mutually perpendicular planes of polarization. The ordinary beam, with the plane of polarization perpendicular to the plane containing the optical axis and the incident beam, obeys the Snell law of refraction at the surface of the prism and travels along the direction of the initial ray path. While the extraordinary beam with vibrations lying in the above plane does not obey the Snell law, but travels in the prism with a speed that varies with direction. Thus, the incident light beam emerges from the prism as two spatially separated beams. Unlike the star which is a point source, the background sky that acts as an extended object illuminates the entire top surface of the prism, hence, producing two broad emergent beams whose centers are spatially separated by an equal amount as above. A considerable amount of overlap between these two beams occurs about the geometrical axis of the prism; therefore, the overlapping region in the focal plane of the telescope remains unpolarized by the prism and provides the background sky (unmodulated) brightness directly. While in the case of the star, the emergent beams from the prism do not overlap. The two images of the star with mutually perpendicular planes of polarization at the focal plane superposed on the unpolarized background sky are observed. Two identical apertures are used to isolate these images, and a rotating chopper is used to alternately block one of the images in order to allow the other to be detected by the same photomultiplier tube. The advantages of such a system are

- the contribution of background sky polarization is eliminated from the data,
- the effect of any time-dependent variations in the sensitivity of the detector is negligible, since the same photomultiplier tube is used as a detector, and

© 2015 by Taylor & Francis Group, LLC

- the quasi-simultaneous detection of both the beams using a fast rotating chopper eliminates the effects of variations in sky-transparency and reduces the errors due to atmospheric scintillation for bright stars significantly; scintillation noise is independent of the brightness of the star and dominates the photon noise for bright objects[88].

The efficiency of a polarimeter at the telescope is determined by

- the faintest magnitude that could be reached with a specified accuracy in a given time and
- obtaining the maximum amount of information on wavelength dependence during the same time.

8.5 Exercises

1. Prove that the Fourier transform (FT) of convolution of two functions is equal to the product of the FTs of those two functions.
2. Prove the following properties of convolution:
   (a) Commutative rule: $f \star g = g \star f$, where $\star$ stands for the convolution parameter
   (b) Associative rule: $f \star (g \star h) = (f \star g) \star h$
   (c) Distributive rule for addition: $f \star (g + h) = f \star g + f \star h$
   (d) If $f = g \star h$, show that $f \star f = (g \star g) \star (h \star h)$
3. Derive the Parseval relation $\int_{-\infty}^{\infty} f(t)g^*(t)dt = \int_{-\infty}^{\infty} \widehat{f}(\omega)\widehat{g}^*(\omega)d\omega$ in which $\widehat{f}(\omega)$ and $\widehat{g}(\omega)$ are the Fourier transforms of $f(t)$ and $g(t)$, respectively.
4. Two polarizing sheets have their polarizing direction parallel so that the intensity, S_0, of the transmitted light is a maximum. Through what angle must either sheet be rotated if the intensity is to drop by one fourth?
5. Imagine x (and y) components of the electric field (plane wave) of a polychromatic beam propagating in z direction as given by equation (8.116). If this wave is passing through a linear retarder followed by a linear analyzer, the amplitude of the electric vector turns out to be

$$E_\phi(\phi, \delta) = E_x \cos\phi + E_y e^{i\delta} \sin\phi$$

   in which δ is the retardance and ϕ the angle between the optical axis of

© 2015 by Taylor & Francis Group, LLC

the analyzer and the x axis. In terms of Stokes vectors that are defined in equation (8.119), derive the Stokes intensity formula.

6. Consider a polarimeter consisting of a quarter-wave plate (QWP) having $\delta = \pi/2$, followed by a linear polarizer (LP) inclined with its optical-axis at 45° to that of the QWP. Mueller matrices of linear polarizer and a rotated retarder are respectively given by equation (8.204) and equation (8.196). Show that this system completely extinguishes left-handed circular polarization (LCP).

7. What are Stokes parameters? Briefly mention their physical significance.

8. Derive the Malus law.

9. If an unpolarized light is perpendicularly incident on two polaroid sheets that are placed on top of each other, what factor is the intensity of the emerging light reduced with respect to the incident light for the directions of the sheets differing by an angle of

   (a) 30°

   (b) 45°

   (c) 60°

10. Describe an astronomical polarimeter.

© 2015 by Taylor & Francis Group, LLC

Chapter 9

Plasma Waves

9.1 What Is a Plasma? 415
9.2 Plasma Parameters 417
9.2.1 Debye Shielding 417
9.2.2 Plasma frequency 418
9.2.3 Gyro frequency 419
9.2.4 Collision Frequency 419
9.2.5 Plasma as a Dielectric 420
9.3 Electrostatic Waves in Magnetized Plasma 420
9.3.1 Ion Cyclotron Waves 421
9.3.2 Lower Hybrid Frequency 422
9.4 Waves in a Cold Plasma 422
9.4.1 Electromagnetic Waves in a Nonmagnetized Plasma 422
9.4.2 Application 423
9.4.3 Electromagnetic Waves in a Magnetized Plasma 424
9.4.4 Case 1: Wave Propagation Perpendicular to the Magnetic Field 424
9.4.5 Case 2: Wave Propagation Parallel to the Magnetic Field . 426
9.5 Plasma Waves (Warm)−Langmuir Waves 429
9.6 Ion-Acoustic Waves 430
9.7 Waves in Nonhomogeneous Plasmas 432
9.8 Quasi-Linear Theory for Nonhomogeneous Plasmas 438
9.8.1 Quasi-Linear Theory 439
9.9 Nonlinear Waves in Plasmas 443
9.9.1 Pondermotive Force 444
9.9.2 Ion-Acoustic Solitons 445
9.9.3 Nonlinear Schrödinger Equation 447
9.9.4 Zakharaov−Shabat Equation 448
9.10 Exercises 449

9.1 What Is a Plasma?

In our daily life, we come across different states of matter. In particular, the most common states of matter are solids, liquids, and gases, in order of decreasing density. If one talks to an astronomer, he will say that only about 5 percent of matter in the universe is in the form of visible matter (the rest being made up of dark matter and dark energy). Of the 5 percent visible matter, more than 90 percent of matter is in the form of the fourth state of matter, plasma. The plasma, although it is prevalent in the universe, is not easily detectable. The plasma can be created in the laboratory under very special

© 2015 by Taylor & Francis Group, LLC

conditions. Most of the stars, our Sun is an ideal, are the natural laboratory of a plasma.

A plasma is an ensemble of an enormous number of moving ions and electrons interacting with each other. In order to describe the behavior of such an ensemble, a distribution function has to be introduced. Boltzmann's and Vlasov's equations will be derived with respect to the distribution function. A plasma viewed as an ensemble of a large number of particles has a large number of degrees of freedom; thus, the mathematical description of plasma behavior is feasible only for simplified analytical models.

Another approach is that of statistical averages in velocity space, such as mass density, flow velocity, and pressure. The magnetohydrodynamic equations for these averages are to be derived. We then obtain a mathematical description of the magnetohydrodynamic fluid model, from which we can study the equilibrium conditions, stability problems, etc., for such a model. Since the fluid model considers only average quantities in velocity space, it is not capable of describing instabilities or damping phenomena, in which the profile of the distribution function plays a significant role. The phenomena which can be handled by means of the fluid model are of low frequency (less than the ion or electron cyclotron frequency); high-frequency phenomena are not describable in terms of it. In Chapters 5 and 6, (homogeneous and non-homogeneous media) we studied waves in MHD (magnetohydrodynamics), which is an approximation of the general description of the plasma.

In a plasma, a large variety of waves can exist. A simple phenomenological classification as transversal and longitudinal waves is insufficient. Instead, we have to consider the conditions under which certain types of waves can exist. For example, in a cold plasma, the thermal motion of the particles and therefore the pressure vanishes. Also the electron temperature is much larger than the gas temperature. Thus, elastic waves cannot exist; they can form only in a warm plasma where a pressure gradient can build up. In an isotropic plasma, no magnetic field exists, which allows modes of propagation other than an anisotropic plasma. Moreover, we also have to consider particle species which are in motion.

In this chapter, we will focus on a model which allows us to study wave phenomena while retaining the essential features of plasma dynamics, while at the same time maintaining relative simplicity in its mathematical form. Such a model is given by a homogeneous plasma of ions and electrons at very low temperatures in a uniform magnetic field. In the unperturbed state, both the ions and electrons of this plasma are motionless. Any small deviation from the unperturbed state induces an electric field and a time-dependent component of the magnetic field, and consequently movements of ions and electrons are excited. The movements of the charged particles induce electric and magnetic fields which are themselves consistent with the previously induced small perturbations. We will use such a concept in this chapter to derive the dispersion relation which characterizes wave phenomena in the cold plasma. Although this model assumes uniformity of the magnetic field, and

© 2015 by Taylor & Francis Group, LLC

the density and also the zero temperature, this cold plasma model is applicable for a nonuniform, warm plasma, if the typical length of variation of the magnetic field and the density is much larger than the wavelength and the phase velocity of the wave is much larger than the thermal velocity of the particles. It is possible to consider the plasma as a medium of electromagnetic wave propagation with a dielectric tensor. This dielectric tensor is a function of the magnetic field and the density which may change with the position. Accordingly plasmas are in general a nonuniform, anisotropic, and dispersive media. When the temperature of the plasma is finite and the thermal velocity of the particles is comparable to the phase velocity of propagating wave, the interaction of the particles and the wave becomes important. Some of the standard texts on plasma physics, dealing with plasma waves are Stix[168] , Cairns[111], Goldston and Rutherford[126], Baumjohann and Treumann[102], Swanson[170], Sturrock[169].

In the next section, we shall define certain important plasma parameters before describing the different types of modes possible in a plasma. We shall also discuss waves in nonhomogeneous plasmas, quasi-linear theory and nonlinear aspects of these waves towards the end of the chapter.

9.2 Plasma Parameters

9.2.1 Debye Shielding

As mentioned earlier, a plasma is an ionized gas composed of roughly equal numbers of electrons and ions (quasi-neutral). To keep things simple, we will be considering a fully ionized gas. Consider a test particle with a positive charge q_T. In a vacuum, the electric potential for such a test particle will be

$$\phi = \frac{q_T}{r} \tag{9.1}$$

If we now introduce the test charge into a uniform plasma of infinite extent, it will repel the surrounding protons while attracting the electrons. Collectively, they form a shielding cloud of particles. If we allow the system to reach a state of equilibrium, then the electric potential due to the test charge in the plasma becomes

$$\phi = \frac{q_T}{r} e^{-r/\lambda_D} \tag{9.2}$$

The potential of the test charge decreases much more rapidly in a plasma than it does in vacuum due to the shielding influence of the charge cloud. For distances $r \gg \lambda_D$, the Debye length, the influence of the test charge is negligible. This phenomenon is generally referred to as the Debye shielding. The Debye length can be calculated and shown to be

$$\lambda_D = \left(\frac{k_B T}{4\pi n e^2} \right)^{1/2} \tag{9.3}$$

© 2015 by Taylor & Francis Group, LLC

where T is the temperature (taken here to be the same for both electrons and protons), n is the electron (or proton) number density, and e is the electric charge. A simple formula for calculating the Debye length is given by $\lambda_D \approx 5\sqrt{T/n}$. In the case of the solar wind, where the $T \approx 10^5$ K and $n \approx 5\ cm^{-3}$, the value of the Debye length is given by $\lambda \approx 7$ m. An alternate definition for a plasma is that it contains a large number of particles in the Debye sphere, i.e., $\Lambda = (4\pi/3)n\lambda_D^3 \gg 1$.

9.2.2 Plasma frequency

Consider a uniform plasma slab of thickness L, in the x direction with very large dimensions in the y and z directions. We shall assume that the plasma has an equal number of electrons and protons. We also assume that the proton mass is very large as compared to the electron mass. This would imply that the protons are effectively fixed in space. Let us displace the electrons from the protons by a distance $\delta \ll L$. This results in an electric field which would exert a force on the electrons, pulling them towards the protons. Allowing the electrons to go, would rush back toward the protons and overshoot, resulting in an oscillation with a typical characteristic frequency.

In order to derive an expression for the frequency, let us use Poisson's equation in one dimension, namely,

$$\frac{\partial E}{\partial x} = 4\pi\rho \tag{9.4}$$

If ρ is the displacement, then $E = 4\pi ne\delta$. The charge per unit area is just $-neL$, while the force per unit area is given by (the electric field) times (the charge per unit area): $F = m_e(d^2\delta/dt^2) = (-neL)(4\pi ne\delta) = -4\pi n^2e^2\delta L$. But the force per unit area is nothing but the mass per unit area times acceleration. Thus the mass per unit area is nM_eL, so we can write

$$nm_eL\left(\frac{d^2\delta}{dt^2}\right) = 4\pi n^2e^2L\delta \rightarrow \left(\frac{d^2\delta}{dt^2}\right) + \frac{4\pi ne^2}{m_e}\delta = 0 \tag{9.5}$$

The above equation represents a harmonic oscillator with a frequency $\omega_{pe} = (4\pi ne^2/m_e)^{1/2}$, called the electron plasma frequency. A similar frequency corresponding to the protons (in general for ions) can also be defined. This frequency, in general, is smaller than the electron plasma frequency. In a thermal plasma, if we take the product of the Debye length with the plasma frequency, we get

$$\lambda_D\omega_{pe} = \frac{(k_BT/4\pi ne^2)^{1/2}}{(4\pi ne^2/m_e)^{1/2}} = \sqrt{\frac{k_BT}{m_e}} = v_{th} \tag{9.6}$$

The cyclic plasma frequency is given by $\nu_{pe} = \sqrt{ne^2/\pi m_e} \approx 9n_e^{1/2}$ k Hz.

© 2015 by Taylor & Francis Group, LLC

9.2.3 Gyro frequency

Consider the motion of electrons and protons in the presence of a magnetic field. The Lorentz force (see Section 2.2.5) is given by

$$\mathbf{F} = m\dot{\mathbf{v}} = \frac{q}{c}\mathbf{v} \times \mathbf{B} \tag{9.7}$$

To begin with, for simplicity let us assume that $\mathbf{B} = B_0\hat{z}$. The equations of motion for the electrons may be written as

$$\begin{aligned} \dot{v_x} &= -\omega_{B_e} v_y & (9.8)\\ \dot{v_{uy}} &= \omega_{B_e} v_x & (9.9)\\ \dot{v_z} &= 0 & (9.10) \end{aligned}$$

Here, $\omega_{B_e} = eB/m_e c$ is the electron gyro frequency. Assuming ϕ to be an arbitrary phase, the solutions can be written as

$$\begin{aligned} \dot{v_x} &= v_\perp \cos(\omega_{B_e} t + \phi)\\ \dot{v_y} &= v_\perp \sin(\omega_{B_e} t + \phi) & (9.11) \end{aligned}$$

where $v_z = v_\parallel = constant$ and $v_\perp = (v_x^2 + v_y^2)^{1/2} = constant$. Integration of the above yields

$$\begin{aligned} v_x &= x_g + \frac{v_\perp}{\omega_{B_e}} \cos(\omega_{B_e} t + \phi)\\ v_y &= v_g + \frac{v_\perp}{\omega_{B_e}} \sin(\omega_{B_e} t + \phi) & (9.12) \end{aligned}$$

The electron's trajectory is a helix, with uniform speed parallel to the magnetic field and circular motion perpendicular to the magnetic field. The radius of the circle is the Larmour radius $r_L = v_\perp/\omega_{B_e}$. Taking $v_\perp \approx v_{th}$, we have the analogy to the product $\lambda_D \omega_p$ that $r_L \omega_{B_e} = v_{th}$. The cyclic gyro frequency is given by

$$\Omega_{B_e} = \frac{eB}{2\pi m_e c} \tag{9.13}$$

It is approximately equal to $2.8B$ MHz, where B, the magnetic field, is in terms of Gauss.

9.2.4 Collision Frequency

We shall briefly describe the collision frequency as its original derivation is rather involved. If the collision, time is defined to be the time required for a given particle to become deviated from its initial trajectory by 90°, the cumulative effects of small-angle collisions far outweigh the occasional large-angle collision. The electron-proton collision frequency for large−angle collision is of the order

© 2015 by Taylor & Francis Group, LLC

$$\nu_L \approx \frac{4\pi n e^4}{m_e^2 v^3} \tag{9.14}$$

where v is the electron speed. However, in contrast, the collision rate due to the cumulative effects of small angle collisions is given by

$$\nu_c = \frac{8\pi n e^4 ln\Lambda}{m_e^2 v^3} \tag{9.15}$$

The electron-proton collision frequency ν_c due to small angle collisions is a factor $2ln\Lambda$ larger than ν_L and the collision time is therefore correspondingly shorter. In both cases, the collision frequency is proportional to $1/v^3$. The kinetic energy of the particle is given by $E = mv^2/2$, so that $\nu_c \propto E^{-3/2}$.

9.2.5 Plasma as a Dielectric

If an electric field is applied to a dielectric, then it induces a polarization $\mathbf{P}$, defined as the induced dipole moment per unit volume:

$$\mathbf{P} = \int d\mathbf{r}\rho(\mathbf{r})\mathbf{r} \tag{9.16}$$

In the case of a uniform plasma, wherein if n electrons per unit volume are moved to a distance $\mathbf{r}$, the equivalent dipole moment per unit volume is nothing but $\mathbf{P} = ne\mathbf{r}$. We know by definition that $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$. If $\mathbf{P}$ is a linear function of the electric field $\mathbf{E}$, such that $\mathbf{P} = \chi\mathbf{E}$, then we can write $\mathbf{D} = \epsilon\mathbf{E}$, where $\epsilon = 1 + 4\pi\chi$ is the dielectric constant of the medium.

If we consider the response of a plasma (nonmagnetized) to an electromagnetic wave with an electric field $\mathbf{E} \approx exp(i\omega t)$, the equation of motion of a single electron is then written as

$$m\frac{d^2\mathbf{r}}{dt^2} = e\mathbf{E} = -m\omega^2\mathbf{r} \tag{9.17}$$

so that $\mathbf{r} = (-e/m\omega^2)\mathbf{E}$. From this expression and that of $\mathbf{P}$, we can show that $\mathbf{P} = (-ne^2/m\omega^2)\mathbf{E}$. From this, one can simplify to show that $\chi = -ne^2/m\omega^2$. Thus, for an unmagnitized plasma, the dielectric constant is nothing but

$$\epsilon = 1 - \frac{4\pi n e^2}{m\omega^2} = 1 - \frac{\omega_{pe}^2}{\omega^2} \tag{9.18}$$

The response of the plasma is embodied in the dielectric constant.

9.3 Electrostatic Waves in Magnetized Plasma

The electrostatic waves in a magnetized plasma, can be divided into waves with $\mathbf{k}$ parallel or perpendicular to $\mathbf{B_0}$. The terms longitudinal or transversal

© 2015 by Taylor & Francis Group, LLC

refer to the direction of the wave vector **k** relative to the fluctuating electric field $\mathbf{E_1}$. Only longitudinal waves are electrostatic as

$$\mathbf{k} \times \mathbf{E_1} = \omega \mathbf{B_1}$$

vanishes. In a transverse wave, $\mathbf{B_1}$ is finite and the wave is an electromagnetic wave. Waves which propagate obliquely to the field can be regarded as the superposition of longitudinal and transverse waves. As for electron oscillations in an unmagnatized plasma in the upper hybrid oscillations, the ions stay fixed in space, creating a positively charged uniform background. The equation of motion contains only the forces exerted by the electric and magnetic fields. For longitudinal waves, the dispersion relation reads

$$\omega_{uh}^2 = \omega_{pe}^2 + \omega_{ce}^2$$

with ω_{pe} being the frequency of electron plasma oscillations and ω_{ce} the electron cyclotron frequency. Only electrostatic waves perpendicular to **B** have this upper hybrid frequency. Disturbances parallel to **B** oscillate with the plasma frequency ω_{pe}; particles moving parallel to the magnetic field do not gyrate and so ω_{ce} vanishes in this case. We can understand these waves as the superposition of two motions. In the plane wave, the electrons exhibit regions of compression and rarefaction as in ordinary plasma oscillations. The magnetic field, which is perpendicular to the direction of the electron motion, forces the electrons into elliptical orbits instead of oscillations along a straight line. As in simple plasma oscillations, the electric field accelerates the electrons, displaced from their rest positions. As the speed of the electrons increases, the Lorentz force exerted by the magnetic field increases, too, reversing the direction of the electron motion. The electrons move against the electric field and, as a result, lose energy. There are two restoring forces acting on the electrons; the electrostatic force resulting from the electron displacement and the Lorentz force. This additional restoring force leads to a higher frequency than that in simple plasma oscillations.

9.3.1 Ion Cyclotron Waves

The upper hybrid wave is a high-frequency wave with ω much larger than both the plasma and the cyclotron frequencies. It results from the motion of electrons in a magnetized plasma. Electrostatic ion waves, like ion acoustic waves, are low frequency waves. Consider an ion acoustic wave with **k** nearly perpendicular to **B**. The force due to the electromagnetic field is the only force considered in the equations of motion. The dispersion relation can be calculated to yield

$$\omega_{pi}^2 = \omega_{ci}^2 + k^2 v_s^2 \tag{9.19}$$

where ω_{ci} is the ion cyclotron frequency. An electrostatic ion wave parallel to

© 2015 by Taylor & Francis Group, LLC

the magnetic field has the same frequency as that of $\omega^2 = k^2 v_s^2$ ion acoustic waves, as the ions do not gyrate around the field and the cyclotron ion frequency vanishes.

9.3.2 Lower Hybrid Frequency

For the ion cyclotron wave, we had assumed $\mathbf{k}$ to be nearly perpendicular to $\mathbf{B}$. The remaining (small) component of the wave vector and the resulting particle motion parallel to $\mathbf{B}$ are essential because they allow the electrons to travel freely along $\mathbf{B}$ to regain thermal equilibrium as is essential for the case of the ion acoustic wave. The relation in this case is

$$\omega_{lh} = \sqrt{\omega_{ce} \cdot \omega_{ci}} \tag{9.20}$$

The motion is maintained in the perpendicular direction only. Thus, we can maintain charge neutrality in that direction. The ions move along $\mathbf{E_1}$ while the Lorentz force acting on them is rather small. The ion displacement along $\mathbf{E_1}$ is limited because the electric field oscillates; the maximum displacement is about $\triangle x_i = eE_1/m_i\omega_{ci}^2$. The electrons gyrate around the magnetic field. In addition to this, they also experience an $\mathbf{E} \times \mathbf{B}$ drift perpendicular to both fields. These waves are of great importance in the auroral regions and are responsible for ion heating.

9.4 Waves in a Cold Plasma

9.4.1 Electromagnetic Waves in a Nonmagnetized Plasma

Let us briefly discuss the properties of electromagnetic waves in an unmagnetized plasma, before going on to the discuss on plasma oscillations. The starting point for this discussion is the Maxwell equations (2.160, 2.175, 2.179, and 2.181; see Section 2.4), which are given by

$$\nabla \cdot \mathbf{D} = 0 \tag{9.21}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{9.22}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{9.23}$$

$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} \tag{9.24}$$

By simple algebra and the fact that $\mathbf{D} = \epsilon \mathbf{E}$, the above set of equations can be reduced to a single wave equation in $\mathbf{E}$ as follows:

$$\nabla^2 \mathbf{E} + \frac{\epsilon}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \tag{9.25}$$

© 2015 by Taylor & Francis Group, LLC

Assuming plane wave solutions of the above equation, the following dispersion relation can be obtained:

$$k^2 = \frac{\omega^2}{c^2}\left(1 - \frac{\omega_{pe}^2}{\omega^2}\right) \rightarrow \omega^2 = \omega_{pe}^2 + k^2c^2 \tag{9.26}$$

Writing

$$\epsilon = k^2c^2/\omega^2 = \mu^2$$

where μ is the refractive index, the group velocity can be written as

$$v_{gr} = \partial\omega/\partial k = \mu c,$$

while the phase velocity is given by $v_{ph} = \omega/k = c/\mu$. For a plasma, $\mu < 1$ so that $v_{gr} < c$ and $v_{ph} > c$. Moreover, if $\omega < \omega_{pe}$, then $\epsilon < 0$, so that μ is imaginary. This implies that the electromagnetic waves are nonpropagating when their frequency is less than the plasma frequency. For $\omega < \omega_{pe}$, an imaginary refractive index results. For a real frequency, an imaginary wave vector **k** would result. Such a wave would not propagate but decay. The second solution, $-ik$, can be interpreted as wave growth. Physically, this is not possible, as we are discussing cold plasma, which is stationary. The energy required for the wave growth cannot be drawn from the plasma. The cold plasma can only absorb the energy of a decaying wave and cannot support wave growth.

9.4.2 Application

Electromagnetic waves can be used for plasma diagnostics in space or in the ionosphere. For example, the density of a space plasma can be determined by detecting the radio signal of a satellite from another satellite or a ground based station. As the frequency of the radio changes, absorption sets in at the plasma frequency of the medium. This will enable us to determine the density. Another interesting application is the measurement of the electron density in the Earth's ionosphere. Radio waves of different frequency are sent from one transmitter to another receiver. As the waves reach the ionosphere, they are refracted (Snell's law). Thus the travel time of a radio pulse between the transmitter and the receiver can be used to determine the height at which the ray path is bent towards Earth. From this geometry, we can also determine the angle at which the electromagnetic wave enters the ionosphere, and with Snell's law, the refractive index can be accurately calculated. The latter gives us the plasma frequency as well as the electron density. Ionospheric reflection is used in short-wave communications as it allows us to send signals around the Earth. In principle, the process can be applied when the vertical geometries of the transmitter and the receiver are located in the same position. Total reflection of the pulse then occurs when the emitted frequency equals the local plasma frequency, i.e., the refractive index is equal to zero. The travel time of the signal gives the height of reflection. Emitting at different frequencies over

© 2015 by Taylor & Francis Group, LLC

a broad frequency band, the height profile of the electrons in the ionosphere can be determined. In reality, however, the situation becomes more complex because the electromagnetic wave propagates into a magnetized plasma and is split into ordinary and extraordinary modes, both having different propagation speeds. We will have more to say about these modes in subsequent sections.

9.4.3 Electromagnetic Waves in a Magnetized Plasma

The magnetized plasma is a very complex medium. In principle, it can support many more modes, in addition to the electromagnetic modes. It can also support a large variety of MHD modes as well. In this discussion, we shall touch upon two special cases of waves in a magnetized plasma. In both cases, we shall restrict ourselves to electron motion, assuming that the ions are fixed, i.e., a cold plasma.

9.4.4 Case 1: Wave Propagation Perpendicular to the Magnetic Field

Consider a plasma permeated by a constant magnetic field $\mathbf{B} = B_0\hat{\mathbf{z}}$. In general, for a perturbation in electro magnetics, the force on an electron is given by the Lorentz force so that the electron equation of motion is given by

$$F = m_e n\dot{\mathbf{v}} = -en\mathbf{E} - \frac{e}{c}n\mathbf{v} \times \mathbf{B} \tag{9.27}$$

If one considers electromagnetic (EM) waves propagating perpendicular to $\mathbf{B}$, then there are two possibilities. The electric field can be parallel to $\mathbf{B}$ or it can be in the x−y plane, perpendicular to $\mathbf{B}$. In the first case, the electric field induces electron motion parallel to $\mathbf{B}$. In this case, the $v \times B$ term vanishes and the problem is exactly the same as that of an unmagnetized plasma. The dispersion relation for such a mode is therefore the same as that of an EM wave in an unmagnetized plasma, i.e., this mode in the literature is often referred to as the $o-$ mode:

$$\omega^2 = \omega_{pe}^2 + k^2c^2 \tag{9.28}$$

The propagation of the ordinary or $o-$ modes is not influenced by the magnetic field; in an ordinary wave , the fluctuating electric field is parallel to $\mathbf{B_0}$, and therefore, the magnetic field does not influence the wave dynamics. Let us look at the second possibility. The electric field induces motion in the x−y plane and the $\mathbf{v} \times \mathbf{B}$ force then causes another velocity component in the x−y plane. The solutions for the $o-$ mode and $x-$ mode are given in Figure 9.1. It is important to realize at this stage that the electron motion of the ordinary mode and this extraordinary mode wave are orthogonal.

© 2015 by Taylor & Francis Group, LLC

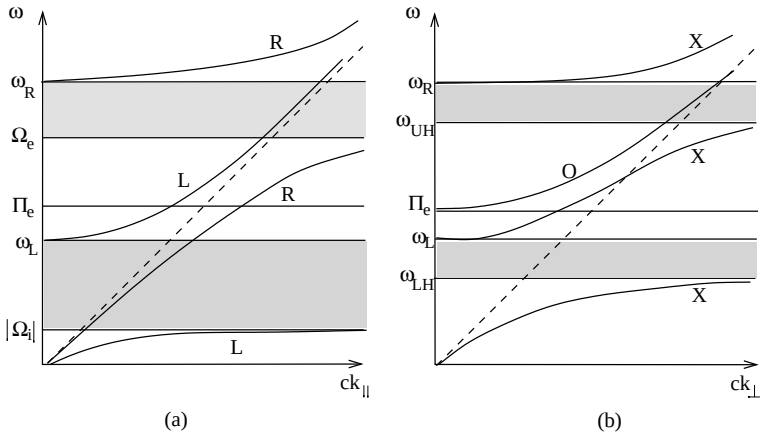


FIGURE 9.1: Solutions for the (a) R and L waves (parallel magnetic field), and (b) O and X waves (perpendicular magnetic field), from Miyomoto[20].

Assuming plane wave solutions, we have

$$-i\omega m_e v_x = -eE_x - \frac{e}{c} v_y B_0 \tag{9.29}$$

$$-i\omega m_e v_y = -eE_y + \frac{e}{c} v_x B_0 \tag{9.30}$$

The expressions for $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{B}$, from Maxwell's equations, can be written as

$$v_x = \frac{-i\omega}{4\pi ne} E_x, \qquad v_y = \left(\frac{ik^2c^2}{4\pi ne\omega} + \frac{-i\omega}{4\pi ne} \right) E_y \tag{9.31}$$

Substituting into the expressions above, one can write, in matrix notation, the relationship as

$$\begin{bmatrix} -i\omega m_e \left(\frac{-i\omega}{4\pi ne} \right) + e & \frac{\epsilon}{c} B_0 \left(\frac{ik^2c^2}{4\pi ne\omega} + \frac{-i\omega}{4\pi ne} \right) \\ -\frac{eB_0}{c} \left(\frac{-i\omega}{4\pi ne} \right) & -i\omega m \left(\frac{ik^2c^2}{4\pi ne\omega} + \frac{-i\omega}{4\pi ne} + e \right) \end{bmatrix} \times \begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The determinant of the coefficients must vanish in order to derive the dispersion relation. After some simple algebra, the dispersion relation for the extraordinary mode can be written as

$$\left(1 - \frac{\omega^2}{\omega_{pe}^2} \right) \left(1 + \frac{k^2c^2}{\omega_{pe}^2} - \frac{\omega^2}{\omega_{pe}^2} \right) + \frac{\omega_{Be}^2 k^2 c^2}{\omega_{pe}^2} - \frac{\omega^2 \omega_{Be}^2}{\omega_{pe}^2} = 0 \tag{9.32}$$

Extracting the factor $k^2c^2/\omega^2 = \mu_\nu^2$ and algebraic simplification yields

$$\mu^2 = \frac{k^2c^2}{\omega^2} = 1 - \frac{\omega_{pe}^2}{\omega^2} \frac{\omega^2 - \omega_{pe}^2}{\omega^2 - \omega_{UH}^2} \tag{9.33}$$

© 2015 by Taylor & Francis Group, LLC

A new frequency called the upper hybrid frequency has been introduced: $\omega_{UH}^2 = \omega_{pe}^2 + \omega_{Be}^2$. The dispersion relation for the $x-$ mode is very complex. It has two important properties, namely, the presence of resonances and cutoffs. A resonance occurs at a given frequency when $k \to \pm\infty$, while a cutoff occurs when at a given frequency $k \to 0$. The dispersion relation reveals that the resonances occur at $\omega = 0$ and $\omega = \omega_{UH}$. The cutoffs are calculated by setting $k = 0$ and then solving for ω. This results in the following relation:

$$\omega = \left[\omega_{pe}^2 + \frac{\omega_{Be}^2}{2} \pm \omega_{Be}\sqrt{\omega_{pe}^2 + \omega_{Be}^2/4}\right]^{1/2} \tag{9.34}$$

An important observation about the $x-$ mode is that it is partially transverse and partially longitudinal. This implies that the electric field of the EM mode has components both parallel and perpendicular to the direction of propagation. The electric vector traces an ellipse in the x−y plane. In the high frequency limit, both the $o-$ and $x-$ modes are linearly polarized. For lower frequencies the refractive index becomes imaginary and the wave vanishes. The ordinary wave has its cutoff at the electron plasma frequency. The refractive index also can go towards infinity. This happens for very large wavenumbers (or very small wavelengths). The corresponding frequency can be found by letting **k** go to infinity. This occurs when ω goes towards the upper hybrid frequency ω_{UH} in the denominator of equation (9.33). As the extraordinary wave approaches this resonance, its phase and group velocities approach zero and the electromagnetic energy is converted into electrostatic oscillations. Thus, between the upper hybrid frequency and the cutoff frequency, a stop band results, wherein the refractive index is less than 0. Here, the extraordinary wave cannot propagate.

The existence of a stop band has an interesting application. Magnetized plasmas may allow the emission of radio waves. Therefore, all planets emit radio signals. In our solar system, the largest radio source is Jupiter. The radio waves are excited close to the planet where both the electron plasma frequency and the electron cyclotron frequency are large. As the waves propagates outward, their frequency decreases. Extraordinary waves excited at a frequency below the upper hybrid frequency will encounter a stop band where ω_{UH} has decreased below the frequency of the wave. These waves cannot escape from the vicinity of the planet. Planetary radio waves observed at large distances, therefore, have not been excited as extraordinary waves below the local upper hybrid frequency.

9.4.5 Case 2: Wave Propagation Parallel to the Magnetic Field

For waves propagating parallel to **B**, we can write the expression for the components of magnetic field and velocity as

$$-ikB_y = \frac{-4\pi ne}{c}v_x - \frac{i\omega}{c}E_x, \qquad ikB_x = \frac{-4\pi ne}{c}v_y - \frac{i\omega}{c}E_y \tag{9.35}$$

© 2015 by Taylor & Francis Group, LLC

and

$$v_x = -\frac{c}{4\pi n}\left(\frac{-ik^2c}{\omega} + \frac{i\omega}{c}\right)E_x, \qquad v_y = -\frac{c}{4\pi n}\left(\frac{-ik^2c}{\omega} + \frac{i\omega}{c}\right)E_y \quad (9.36)$$

As in the previous case, we can write the above set of equations in a matrix form. Let the determinant of the coefficients to zero and obtain a solution which is given by

$$\left(1 + \frac{k^2c^2}{\omega_{pe}^2} - \frac{\omega^2}{\omega_{pe}^2}\right)^2 = \frac{\omega_{Be}^2}{\omega_{pe}^2}\left(\omega - \frac{k^2c^2}{\omega}\right)^2 \quad (9.37)$$

Taking the square root of the above and extracting the refractive index yields

$$\mu^2 = \frac{k^2c^2}{\omega^2} = 1 - \frac{\omega_{pe}^2/\omega^2}{1 \pm \omega_{Be}/\omega} \quad (9.38)$$

The above expression implies that there are two wave modes corresponding to the two sign in the denominator of the last term. If we take the + sign, the corresponding wave mode will be described as right circularly polarized ($R - wave$) and for the - sign, it will be left circularly polarized ($L - wave$). This implies that the electric field vector of the EM mode traces a circle in the counterclockwise (RCP) or clockwise (LCP) sense. Here again, there are resonances and cutoffs in the dispersion relation. The $R-$ wave has a resonance at $\omega = -\omega_{Be}$. This implies that the frequency of rotation of the electric field matches with the gyro motion of the electrons. However, the $L-$ mode does not have any resonances. The cutoffs can be obtained by letting $k = 0$, which is given by

$$\omega_{R,L} = \mp\frac{\omega_{Be}}{2} + \sqrt{\omega_{pe}^2 + \omega_{Be}^2/4} \quad (9.39)$$

It is easy to realize that in the high frequency limit, the $R-$ and $L-$ mode propagate near c. The solution of the dispersion relation for electromagnetic waves propagating parallel to the magnetic field is shown in Figure 9.2. There are two distinct cases : $\omega_{pe} < \omega_{ce}$ and $\omega_{pe} > \omega_c e$. The main difference between these cases is the cutoff for the left hand mode : for $\omega_{pe} < \omega_{ce}$ the cutoff is below the electron cyclotron frequency, while in the opposite case it is above. For a constant cyclotron frequency, these waves approach a refractive index of value 1, for high frequencies. The $R-$ wave has a stop band between ω_R and ω_{ce}. The $R-$ waves in the lower frequency range are also called Whistler waves and are important for propagation studies in the magnetosphere. Whistler waves can be excited by lightning discharge. Thus, the source has a short duration but creates a wide spectrum of different frequencies. Since the propagation time depends on the group speed, the first waves arriving at a distant observer have higher frequencies than those arriving later, rather like a whistle with decreasing pitch. The rate of frequency change contains information about the change in plasma density along the propagation

© 2015 by Taylor & Francis Group, LLC

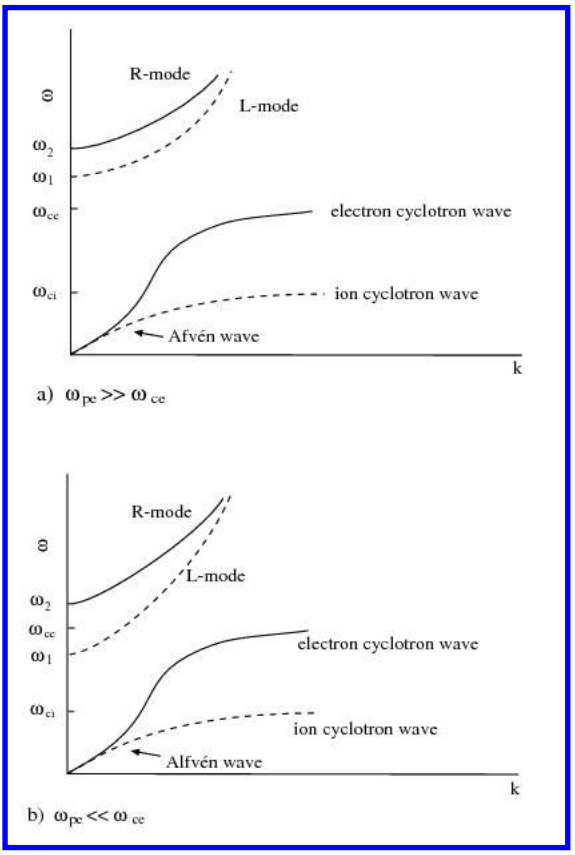


FIGURE 9.2: Parallel propagation for the (a) high plasma density and (b) low plasma density. Cutoffs are found when the dispersion curve meets the frequency axis ($k = 0$) and resonances are found for large values of k, from Miyomoto[20].

path. The solution of the dispersion relation depicting the Whistler mode is shown in Figure 9.3.

An important phenomenon arising from the fact that the $x-$ and the $o-$ modes are described by different dispersion relations is the Faraday rotation . A linearly polarized wave can be regarded as a superposition of two circularly polarized modes. The propagation speeds of the two modes differ and the two circularly polarized modes therefore become increasingly out of phase as they propagate. This results in the plane of polarization rotating as the wave propagates. The change in the phase ϕ of a circularly polarized mode as it travels a distance d is given by $2\pi d/\lambda = \mathbf{k} \cdot \mathbf{d}$. For a medium whose properties

© 2015 by Taylor & Francis Group, LLC

change along d, we have

$$\phi_{R,L} = \int_0^d k_{R,L} ds \tag{9.40}$$

where $k_{R,L} = \omega\sqrt{\epsilon_{R,L}}/c$. With simple algebra, we can show that

$$k_{R,L} \approx \frac{\omega}{c}\left[1 - \frac{\omega_{pe}^2}{2\omega^2}\left(1 \mp \frac{\omega_{Be}}{\omega}\right)\right] \tag{9.41}$$

The angle with which the plane of polarization rotates is given by

$$\triangle\theta = \frac{1}{2}\int_0^d (k_R - k_L) ds = \frac{1}{2}\int_0^d \frac{\omega_{pe}^2 \omega_{Be}}{c\omega^2} = \frac{2\pi e^3}{m_e^2 c^2 \omega^2}\int_0^d nB_\parallel ds \tag{9.42}$$

In passing, we may say that the Faraday rotation is an observable effect and plays an important role in diagnosing a cosmic magnetic field.

9.5 Plasma Waves (Warm)−Langmuir Waves

In the case of plasma oscillations in a cold plasma, the group velocity is zero and the disturbance does not propagate. However, in a warm plasma, the situation is completely different. The thermal motion of the electrons carries information about a disturbance into the undisturbed ambient plasma. This disturbance will propagate in the form of a wave. In order to derive the dispersion relation in this case, one has to simply add the pressure gradient force $-\nabla p$ to the equation of motion. The algebra for calculating the dispersion relation is identical to that for a cold plasma and will not be repeated. For a plasma which is adiabatic, the dispersion relation for the plasma wave can be written as

$$\omega^2 = \omega_{pe}^2 + \frac{3}{2}k^2 v_{th}^2 \tag{9.43}$$

for the one-dimensional case. For three dimensions, it will be

$$\omega^2 = \omega_{pe}^2 + \frac{5}{3}k^2 v_{th}^2 \tag{9.44}$$

where the thermal velocity is defined as $v_{th} = \sqrt{2k_B T/m_e}$.

The above relations are satisfied subject to the assumption that the local equilibrium is Maxwellian. This means that the plasma must allow frequent collisions. However, in space plasmas, in general, this is not true. In this case, the dispersion relation gets modified to yield

$$\omega^2 = \omega_{pe}^2 + 3k^2 v_{th}^2 \tag{9.45}$$

© 2015 by Taylor & Francis Group, LLC

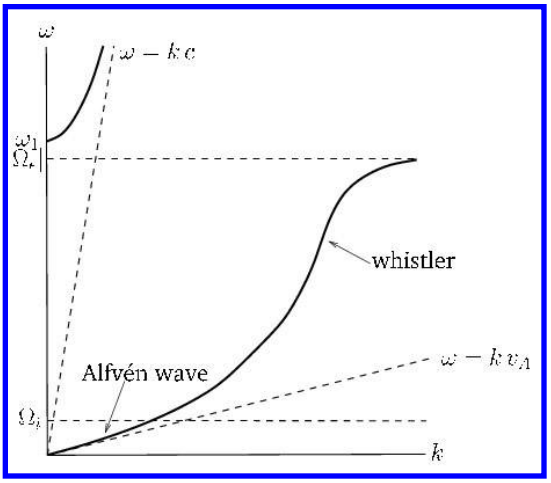


FIGURE 9.3: Solution of the dispersion relation for a right-handed wave propagating parallel to the magnetic field in a magnetized plasma, from Miyomoto[20].

The above relation in the literature is generally called the Bohm−Gross relation. The group velocity of the plasma waves in space can be calculated to yield

$$v_g = \frac{d\omega}{dk} = \frac{3k}{\omega} v_{th}^2 = \frac{3v_{th}^2}{v_{ph}} \tag{9.46}$$

The characteristic of the electrostatic Langmuir wave, or electron plasma wave, is shown in Figure 9.4 (horizontal dot-dashed line at $\omega/\omega_{pe} = 1$). It is easy to realize that the group velocity vanishes as the thermal energy of the plasma approaches zero. This is the case for cold plasma, as mentioned earlier. The group velocity is, in general, smaller than the thermal speed and so much smaller than the speed of light. For large wavenumbers **k**, the information propagates with the thermal velocity. However, for small wavenumbers, the information travels slower than v_{th} as the density gradient decreases for large wavelengths so that the net flux of momentum into the adjacent layers tends to be small.

9.6 Ion-Acoustic Waves

In Langmuir waves, which are plasma oscillations, the ions have much more mass than the electrons. They stay relatively fixed at their position. Thus, one can categorize Langmuir waves as high frequency waves. If we allow motion

© 2015 by Taylor & Francis Group, LLC

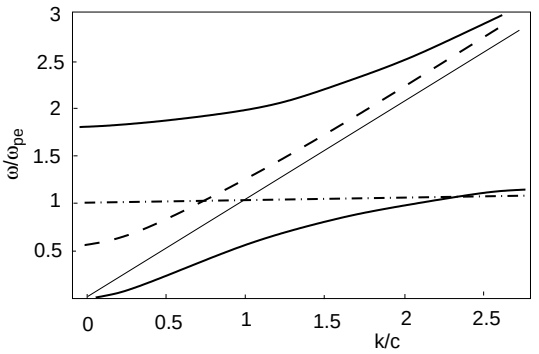


FIGURE 9.4: The electrostatic Langmuir wave (electron plasma wave) is the horizontal dot-dashed line at $\omega/\omega_{pe} = 1$, from Bastian [21].

of the ions, then the situation starts changing. The inertia of the ions requires rather slow oscillations. Thus, ion waves may be categorized as low-frequency waves. The sound waves are analogous to the ion waves in ordinary fluids. The dispersion relation for sound waves can be derived from the Navier-Stokes equations, keeping the pressure gradient force on the right−hand side of the momentum equation. The dispersion relation for an infinite medium can be simply written as $v_s = \omega/k = \sqrt{\gamma p_0/\rho_0} = \sqrt{\gamma k_B T/m}$, with v_s being the speed of sound, γ is the ratio of specific heats, and p_0 and ρ_0 are the pressure and the density (undisturbed), respectively. It is well known that sound waves are pressure waves. They transport momentum from one layer to the next due to collisions between molecules or atoms. In spite of the fact that it has low density, a similar phenomenon arises in a plasma. In this case, the momentum is transported by Coulomb collisions; thus, information is contained in the charges and the fields. As we are considering the motion of both the electrons and the ions, the ion wave can be derived in the framework of a two-fluid theory. In this case, one considers the pressure gradient force as well as the force exerted by the electromagnetic field. In this case, the dispersion relation gets modified as

$$\frac{\omega^2}{k^2} = \frac{\gamma_e k_B T_e + \gamma_i k_B T_i}{m} \approx v_s \tag{9.47}$$

An important note is that $\gamma_e \approx 1$ as the electrons are fast compared with the waves so that an isothermal distribution is established easily. However, the ions experience a one-dimensional compression in the plane so that $\gamma_i \approx 3$. The group velocity of the ion wave is independent of the wavenumber k as is evident from the expression for the square of the phase velocity.

If we assume the ions to be cold, so that $T_i \to 0$ and the wave length is small, then the ions will oscillate with the ion plasma frequency (similar to

© 2015 by Taylor & Francis Group, LLC

the electron plasma frequency) as

$$\omega^2 = \omega_{pi}^2 = \frac{n_i Z^2 e^2}{\epsilon_0 m_i} \tag{9.48}$$

The frequency of the ion acoustic wave asymptotically approaches that of the ion plasma wave for $k\lambda_D \to 3$. It is important at this stage to realize that there are fundamental differences in the dispersion relation for the electron and ion waves. Electron plasma waves are waves with constant frequency, wherein the thermal motion adds only a small correction. However, the ion waves are waves with a constant speed and require thermal motion. In this case, the phase speed and the group speed are identical (phase velocity independent of the wavenumber). The basic difference between these two waves arises due to the behavior of the second particle component (two-fluid description). In the case of an electron plasma wave, the electrons oscillate while the ions stay fixed. For the ion plasma waves, the ions oscillate with the electrons in motion. In space plasmas, the ion acoustic waves are observed in the upstream of planetary bow shocks, where they are generated by super thermal particles streaming away from the shock front.

Table 1 summarizes the different types of waves that are present in a plasma. We have included the MHD modes (Alfven and the magneto−acoustic) for the sake of completion. It is important to realize that the discussion so far has the constraint that it pertains only to linear analysis also for homogeneous plasmas. In what follows, we shall discuss waves in non-homogeneous media as well as touch upon quasi-linear and nonlinear studies pertaining to these waves.

9.7 Waves in Nonhomogeneous Plasmas

The concept of uniform plasmas in the previous sections are situations rarely realized, although it is possible in some bounded regions, where the approximation may be reasonably good. In principle, one can assume a plasma to be uniform if the plasma parameters do not vary much over a wavelength. In this section, we shall concentrate on situations where the wave energy behaves in a slowly varying medium, and what happens in those regions where the theory of geometric optics or WKB (Wentzel-Kramer-Brillouin) breaks down. The analysis in these special regions includes the effects of reflection at cutoffs, and absorption and mode conversion at resonances. We have talked about cutoffs and resonances in the previous sections. Density gradients (finite) may also give rise to electric fields and drift waves, along with new drift and gradient driven instabilities. We will have more to say about instabilities in the final chapter. In many instances, these various effects can be separated

© 2015 by Taylor & Francis Group, LLC

TABLE 9.1: Different Types of Plasma Waves

Wave	Geometry	Dispersion relation
Electron Waves (Electrostatic)		
Langmuir waves	$\mathbf{B_0} = 0$	
	or $\mathbf{k} \| \mathbf{B_0}$	$\omega^2 = \omega_{pe}^2 + 3k^2 v_{th}^2/2$
Upper hybrid waves	$\mathbf{k} \perp \mathbf{B_0}$	$\omega_{UH}^2 = \omega_{ci}^2 + \omega_{ce}^2$
Ion Waves (Electrostatic)		
Ion acoustic waves	$\mathbf{B_0} = 0$	
	or $\mathbf{k} \| \mathbf{B}$	$\omega^2 = k^2 v_s^2$
Ion cyclotron waves	$\mathbf{k} \perp \mathbf{B_0}$	$\omega^2 = \omega_{ci}^2 + k^2 v_s^2$
Lower hybrid waves	$\mathbf{k} \perp \mathbf{B_0}$	$\omega_{lh}^2 = \omega_{ci} \cdot \omega_{ce}$
Electron Waves (Electromagnetic)		
Light waves	$\mathbf{B_0} = 0$	$\omega^2 = \omega_{pe}^2 + k^2 c^2$
O-waves	$\mathbf{k} \perp \mathbf{B_0}$	
	or $\mathbf{E_1} \| \mathbf{B_0}$	$\omega^2 = c^2 k^2 + \omega_{pe}^2$
X-waves	$\mathbf{k} \perp \mathbf{B_0}$	$\omega^2 = c^2 k^2 +$
	or $\mathbf{E_1} \| \mathbf{B_0}$	$\omega_{pe}^2 \cdot (\omega^2 - \omega_{pe}^2)/(\omega^2 - \omega_h^2)$
Whistler (R-waves)	$\mathbf{k} \| \mathbf{B_0}$	$\omega^2 = c^2 k^2 - \omega_{pe}^2/[1 - (\omega_{ce}/\omega)]$
L-waves	$\mathbf{k} \| \mathbf{B_0}$	$\omega^2 = c^2 k^2 + \omega_{pe}^2/[1 + (\omega_{ce}/\omega)]$
Ion Waves (Electromagnetic)		
Alfven waves	$\mathbf{k} \| \mathbf{B_0}$	$\omega^2 = k^2 v_A^2$
Magneto-acoustic waves	$\mathbf{k} \perp \mathbf{B_0}$	$\omega^2 = c^2 k^2 \cdot (v_s^2 + v_A^2)/(c^2 + v_A^2)$

and dealt with one at a time, but there are important cases where a combination of effects occurs, several of which are simultaneously important, and each individual technique breaks down. We will generally discuss these effects in isolation, although the analysis of a cutoff resonance pair usually includes both mode conversion and absorption and will be treated as a single problem. To begin with as an illustration, we shall deal with the WKB method for one-dimensional inhomogeneities.

Consider an unmagnetized plasma which has only variation in the density. The starting point of further analysis is the wave equation which is written as

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} + \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = \mu_0 \frac{\partial \mathbf{J}}{\partial t} \tag{9.49}$$

This equation depends only on variation in the plasma density through the plasma current $\mathbf{J}$. We shall assume the density $n_0 = n_0(x)$. The equilibrium

© 2015 by Taylor & Francis Group, LLC

quantities are independent of y and z, this being a one dimensional problem. We shall assume Fourier and harmonic time dependence so that

$$\nabla \to \frac{d}{dx}\hat{e_x} + ik_y\hat{e_y} + ik_z\hat{e_z}, \quad \text{and} \quad \frac{\partial}{\partial t} \to -i\omega \tag{9.50}$$

With the simplification that the plasma is isotropic, one can rotate the coordinate system about the x axis until the wave has no k_z component. The current may be derived from the equations of motion for ions and electrons given by

$$m_j \frac{\partial v_j}{\partial t} = q_j \mathbf{E}, \qquad j = i, e \tag{9.51}$$

and

$$j = (e)n_0(x)(v_i - v_e) \tag{9.52}$$

with the result that

$$j \approx \frac{(i)n_0(x)e^2}{\omega m_e}\mathbf{E} = \frac{i\epsilon_0}{\omega}\omega_{pe}^2(x)\mathbf{E} \tag{9.53}$$

where $\omega_{pe}^2 = n_0(x)e^2/m_e\epsilon_0$. The z component of the wave equation which is not coupled with the other components can be written as

$$-\left(\frac{d^2}{dx^2} - k_y^2\right)E_z + \left[\frac{\omega_{pe}^2(x)}{c^2} - \frac{\omega^2}{c^2}\right]E_z = 0 \tag{9.54}$$

Writing $y(x) = E_z(x)$ and

$$k^2(x) = \frac{\omega^2 - \omega_{pe}^2(x)}{c^2} - k_y^2 \tag{9.55}$$

the above differential equation can be cast into the WKB form as

$$\frac{d^2y}{dx^2} + k^2(x)y = 0 \tag{9.56}$$

If we assume that $k^2(x)$ is slowly varying, then the WKB method allows us to find good approximations to the exact solutions in rather general form. If for example, $k(x) = constant$, then the solution of the equation is trivial and has the form

$$y = A_1e^{ikx} + A_2e^{-ikx} \tag{9.57}$$

The above solution represents waves traveling to the left and to the right. Looking for solutions which are similar to that of the uniform result, we shall assume a solution of the form

$$y(x) = A(x)e^{i\psi(x)} \tag{9.58}$$

$A(x)$ is assumed to be slowly varying amplitude whereas $\psi(x)$ is a rapidly

© 2015 by Taylor & Francis Group, LLC

varying function such that $\psi'(x) = \pm k(x)$. The first and second derivatives of the above equation are given by

$$y' = ikAe^{i\psi} + A'e^{i\psi}$$

$$y'' = -k^2Ae^{i\psi} + 2ikA'e^{i\psi} + ik'Ae^{i\psi} + A''e^{i\psi}$$

so that equation (9.56) becomes

$$A'' + 2ikA' + ik'A = 0 \tag{9.59}$$

Assuming that A'' is rather small, then to the lowest order $A(x) = 1/\sqrt{k(x)}$, the complete solution can be written as

$$y(x) = \frac{A_0}{\sqrt{k(x)}} exp[\pm i \int^x k(x')dx'] \tag{9.60}$$

Let us assume this to be lowest order solution and introduce a small correction term as follows

$$A(x) = [1 + \eta(x)]/\sqrt{k(x)}$$

where once again we assume that η is a slowly varying function of x and that η'' may be neglected. Inserting the correction into equation (9.59), we have

$$\frac{\eta'}{1+\eta} = \frac{1}{4i}\frac{(k''/k^2 - (3/2)k''^2/k^3)}{1 + ((1/2i) + k'/k^2)} \tag{9.61}$$

so that

$$\eta \approx \frac{1}{4i}\frac{k'}{k^2} \tag{9.62}$$

The validation of the above constraint may be written as $|\eta| \ll 1$ or

$$\left|\frac{1}{k}\frac{dk}{dx}\right| \ll k. \tag{9.63}$$

The constraint implies that the change in the wavelength over a given wavelength is very small. The disadvantage in this approximation is that it fails when $k \to 0$ or when $k' \to \infty$ and also when the wave approaches a cutoff or a resonance. We shall briefly discuss the behaviour of the wave near a cutoff, as it is important to justify further analysis. In the neighborhood of the cutoff, we expand $k^2(x)$ about the cutoff:

$$k^2(x) = k^2(x_0) + \frac{d}{dx}k^2(x)|_{x_0}(x - x_0) + \mathcal{O}(x - x_0)^2$$

where $k^2(x_0) = 0$ defines x_0 and we define the coefficient such that $k^2(x) = \beta^2(x - x_0)$. The result of the solution (equation 9.60) is valid whenever $|x -$

© 2015 by Taylor & Francis Group, LLC

$x_0| \gg \beta^{-2/3}$. The behavior near the cutoff must come from the solution of the differential equation which may be written as

$$\frac{d^2y}{dx^2} + \beta^2(x - x_0)y = 0$$

Introducing the change of variable as $z = -\beta^{2/3}(x - x_0)$, the equation can be written as

$$y'' - \zeta y = 0 \tag{9.64}$$

which is the Airy equation (a special function). The solutions of the Airy equation are very well known (see any book on special functions). It has independent solutions

$$y(\zeta) = C_1 Ai(\zeta) + C_2 Bi(\zeta)$$

Introducing $\xi = (2/3)\zeta^{3/2}$, the asymptotic forms of the Airy functions can be written as

$$Ai(\zeta) = \frac{1}{2\sqrt{\pi}\zeta^{1/4}} e^{-\xi} \tag{9.65}$$

$$Ai(-\zeta) = \frac{1}{\sqrt{\pi}\zeta^{1/4}} [\sin(\xi + \pi/4) - \cos(\xi + \pi/4)] \tag{9.66}$$

$$Bi(\zeta) = \frac{1}{\sqrt{\pi}\zeta^{1/4}} e^{\xi} \tag{9.67}$$

$$Bi(-\zeta) = \frac{1}{\sqrt{\pi}\zeta^{1/4}} [\sin(\xi + \pi/4) + \cos(\xi + \pi/4)] \tag{9.68}$$

These asymptotic solutions need to be matched with the approximate solutions given by equation (9.60) and are given by

$$y(x) = \frac{A_1}{(x - x_0)^{1/4}} e^{i(2/3)\beta(x-x_0)^{3/2}} + \frac{A_2}{(x - x_0)^{1/4}} e^{-i(2/3)\beta(x-x_0)^{3/2}}, \quad x > x_0 \tag{9.69}$$

$$y(x) = \frac{B_1}{|x - x_0|^{1/4}} e^{i(2/3)\beta|x-x_0|^{3/2}} + \frac{B_2}{|x - x_0|^{1/4}} e^{-i(2/3)\beta|x-x_0|^{3/2}}, \quad x < x_0 \tag{9.70}$$

As $\zeta \to \infty (x \to -\infty)$, $Bi(\zeta) \to \infty$ so that in an unbounded plasma, we must require $c_2 = 0$ in the solution of equation (9.64). Writing $\xi = -(2/3)\beta(x - x_0)^{3/2}$, we can write

$$Ai(\zeta) \approx (\frac{2}{\pi\sqrt{\zeta}})^{1/2} \sin\zeta = -\frac{\sqrt{2/\pi}e^{-i\pi/4}}{(x - x_0)^{1/4}\beta^{3/8}} \sin[(2/3)\beta(x - x_0)^{3/2}] \tag{9.71}$$

This implies that we require $A_2 = A_1$. The above result also indicates that there is total reflection, since the amplitudes of the incoming and the outgoing waves are the same. The asymptotic expressions, once far from the cutoff, can be matched onto more realistic expressions for $k(x)$ than the linear ones, and

© 2015 by Taylor & Francis Group, LLC

the phase of the reflected wave at a distant point can be estimated. Whenever the plasma is bounded, either due to the fact that there is a definite plasma edge, or due to the density or field not changing monotonically so that another cutoff is nearby, both $Ai(z)$ and $Bi(z)$ functions must be used to satisfy the boundary conditions or connection formulas. The back-to-back cutoff problem, where two cutoffs occur a finite distance apart and the wave is not propagating between the cutoffs, is treated in the following section. This situation leads to tunneling and only partial reflection. The analysis of resonances is intrinsically more difficult than the analysis of cutoffs, because the physics of what resolves the resonance must be included in order to obtain physically meaningful results. While effects of losses are invariably important near resonances, the effects of collisions in high temperature plasmas frequently are so weak as to be insignificant if any other effects are involved. While collisions may be the only effect in cold plasmas which could resolve resonances, thermal effects are frequently more important than collisions. This was demonstrated by Stix [168] when he investigated the nature of the lower hybrid resonance in a finite temperature plasma. He discovered that rather than approach a true pole as the wave approached the resonance, the $k^2(x)$ for the wave doubled back and continued on a warm backward wave branch, leading to the phenomenon of mode conversion where a wave of one type (mode) is linearly coupled to a wave of another type (mode). The example shown actually has a second mode-conversion point where its mode converts into a forward wave once again. The phenomenon can be understood in terms of Huygen's principle, where the wave can be thought of as shaking the plasma at a particular frequency and wavenumber at each point. If there are two waves that have the same wavenumber and matching phase velocity at that point, the excited plasma will excite both waves and transfer some energy between the waves. It is even possible for oppositely directed group velocities to be coupled, provided the phase velocities are in the same direction, and for all of the wave energy to be transferred from one mode to the other so that the wave energy flow turns around without reflecting where we define reflection as being due to wave energy coming back on the incident branch or mode.

The simplest model which can represent two types of coupled waves traveling in either direction is a fourth-order wave equation (since there are two kinds of waves each having two components traveling in opposite directions), and the isolated resonance may be modeled by the Wasow equation,

$$y^{iv} + \lambda^2(\zeta y'' + \beta y) = 0 \tag{9.72}$$

which was analyzed by Wasow [89] and generalized by Rabenstein [90] to include the first derivative and the complex coefficients. The expression for $k^2(\zeta)$ is given by

$$k^4 - \lambda^2\zeta k^2 + \lambda^2\beta = 0 \tag{9.73}$$

Here λ is a parameter defined in equation (9.72). The asymptotic solutions for $k^2(\zeta) \approx \lambda^2\zeta, \beta/\zeta$. On the contrary, the WKB type of analysis leads to asymptotic forms like $\zeta^{1/4}exp(\pm 2i\sqrt{\beta\zeta})$ and $\zeta^{-5/4}exp(\pm 2i\lambda\zeta^{3/2}/3)$. The difficulty is

© 2015 by Taylor & Francis Group, LLC

that the connections between these is not obvious until the differential equation has been solved exactly and the coupling established by the asymptotic forms of the exact solutions.

The connection formula is given by

$$-(\beta\zeta)^{1/4}e^{-2\sqrt{-\beta\zeta}} \rightarrow \frac{\beta}{\lambda^{3/2}\zeta^{5/4}}e^{-i((2/3)\lambda\zeta^{3/2}+\pi/4)} + i\sqrt{\pi\beta\zeta}H_1^{(2)}(2\sqrt{\beta\zeta}) \tag{9.74}$$

Here the Hankel function varies as $exp(-2i\sqrt{\beta\zeta})$. The conclusion from this connection formula is that the incident cold wave, represented by the Hankel function term, is coupled to the warm wave, represented by the $exp(-2i\lambda\zeta^{3/2}/3)$ term, both on the right-hand side, and that there is no reflection. The fact that both phase terms are negative is indicative of the fact that the phase velocities match at the turning point, but in this case, the group velocities do not as they travel in opposite directions. The scenario is that the wave energy on the cold branch is totally mode converted to the warm branch which propagates back away from the hybrid resonance layer. The exponentially decaying term for $x < 0$ carries no energy on that side unless a boundary is nearby.

Mode conversion theory was developed to deal with resonances in inhomogeneous plasmas, taking into account the influence of the merging of two different types of wave modes along with tunneling, reflection, conversion, and absorption. The method is nontrivial in its mathematical sophistication, but it incorporates a substantial amount of physics and is involved in virtually every type of plasma wave energy absorption except Landau damping and collisional damping. The details will be skipped for brevity.

9.8 Quasi-Linear Theory for Nonhomogeneous Plasmas

In the discussion above, we considered plasma wave phenomena where the treatment was restricted to linear waves, so that particles and waves have a self-consistent motion. However, the effects of finite amplitude waves was neglected. For some externally excited waves, this is adequate; when the amplitude is rather low, however, for unstable waves, the amplitude will eventually become large enough to affect the zero−order solution and couple the primary wave to other waves. A large number of finite amplitude waves may make the plasma turbulent and nonlinear effects may start dominating.

Unlike the hydrodynamic case, where the turbulent state is particularly difficult due to the various nonlinear couplings, where it is rather difficult to isolate the various nonlinear couplings, in a plasma, this may not be the case. The reason being that the nonlinearity induced waves are frequently excited with small amplitudes, unless these waves are also unstable. As a first

© 2015 by Taylor & Francis Group, LLC

approximation, we may consider the different modes to be independent and then add the coupling at various levels and focus on one phenomenon at a time, for the weak turbulence.

Without loss of generality, in the first approximation, we shall assume that the nonlinear coupling coefficients between various linear modes change slowly with time. This we may interpret as the weak turbulence theory. The typical parameter used for expansion is the ratio of the wave energy to the plasma energy density. This theory is, in general, broken into three types of interactions: (1) quasi-linear wave-particle interactions, (2) nonlinear wave-wave interactions, and (3) nonlinear wave-particle-wave interactions. All these situations arise due to resonance conditions which can be viewed as due to conservation of energy and momentum. We shall discuss only a direct perturbation approach in this chapter.

9.8.1 Quasi-Linear Theory

Quasi-linear theory was developed to treat weak turbulence observed in nonequilibrium plasmas. Due to the non-Maxwellian nature of a turbulent plasma, a spectrum of electromagnetic waves is generated from local charge separations or particle currents. This spectrum of waves, known as fluctuations, can then interact with the plasma through the Landau processes. On a short enough time scale, such that $t \ll \tau_B = \sqrt{m/ek^2\phi_0}$ where τ_B is the bounce time for a particle in the electrostatic potential well of the wave. For the discussions that follow, the linear Landau theory is adequate. However, the kinetic energy gain of the particles is proportional to the slope of the zero-order distribution function, $\partial f_0/\partial v|_{v=v_p}$. This implies that the damping or growth rate of the wave is independent of time even on the longer time scale over which the first-order distribution function changes since the zero-order distribution is unaffected by the growth of the first-order distribution function. At some time, this clearly fails to describe the physics of wave-particle interaction if the first order distribution function grows to be a non-negligible perturbation on the zero−order distribution. It is the purpose of quasi-linear theory to address the time evolution of the zero-order distribution function f_0. Moreover, we will show that this theory conserves energy and momentum between the wave and the particle distribution function. Another useful aspect of quasi-linear theory is the interaction of the particles with an externally imposed wave such as in wave heating of a plasma or driving a macroscopic current in a plasma with waves. We should be cautious in applying these results, since these wave damping or wave growth processes are collisionless, so the final state of the distribution function will not be Maxwellian. The treatment of only one nonlinear effect at a time for a non-Maxwellian distribution may occasionally lead to a serious error, especially as the amplitude grows, since somewhere one crosses the boundary from weak turbulence to strong turbulence. One rather surprising experimental observation is that the quasi-

© 2015 by Taylor & Francis Group, LLC

linear description of the wave−particle interaction is quite good even for fairly large amplitude waves.

For simplicity, we shall first restrict our discussion to the case of an unmagnetized plasma and develop the more general case later. This will enable us to bring out the physical processes more clearly without the mathematical complexity added by the magnetic field. Furthermore, the one-dimensional problem will be applicable to wave-particle interactions along a magnetic field where the Lorentz force is negligible. The basic procedure of developing the one-dimensional quasi-linear equations is the same as the treatment of the linear Landau case. Here, however, the quasi-linear approximation implies that the distribution function is allowed to evolve nonlinearly while the wave is calculated using the linear theory. The distribution function is assumed to be locally uniform in space, but it approaches zero as one tends toward infinity, or $f(x \to \pm\infty, v, t) = 0$. We again assume there are no zero order electric or magnetic fields and use the Vlasov equations for electrons only (assuming a uniform, stationary ion background to cancel all zero-order electric fields).

The starting point for further discussion is based on the following equation:

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} - \frac{e}{m}E\frac{\partial f}{\partial v} = 0 \tag{9.75}$$

where the distribution function has the following properties:

$$f(x,v,t) = f_0(v,t) + f_1(x,v,t) \tag{9.76}$$

$$f_1(x,v,t) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \bar{f}_k(k,v,t) e^{ikx} \tag{9.77}$$

$$E(x,t) = \int_{-\infty}^{\infty} \frac{dq}{2\pi} \bar{E}_q(q,t) e^{iqt} \tag{9.78}$$

The zero-order distribution functions depends on time and the bar denotes the Fourier-transformed quantity. Denoting the average over space by

$$\langle f \rangle = \lim_{L\to\infty} \frac{1}{L} \int_{-L/2}^{L/2} d\chi f \tag{9.79}$$

the average of the first-order perturbations tend to zero so that

$$\langle f \rangle = f_0(v,t) \tag{9.80}$$

$$\langle E \rangle = 0 \tag{9.81}$$

The average of the second term of equation (9.75) can be eliminated by the condition

$$\left\langle \frac{\partial f}{\partial x} \right\rangle = \lim_{L\to\infty} \frac{1}{L} \int_{L/2}^{L/2} d\chi \frac{\partial f}{\partial x} = 0 \tag{9.82}$$

as $f(x \to \pm\infty) = 0$. Thus, we have a nonlinear equation for the time evolution of the zero-order distribution by averaging equation (9.75) as

© 2015 by Taylor & Francis Group, LLC

$$\frac{\partial}{\partial t} f_0(v,t) = \frac{e}{m_e}\frac{\partial}{\partial v}\langle E f_1\rangle \tag{9.83}$$

where

$$\langle E f_1\rangle = \lim_{L\to\infty}\frac{1}{L}\int_{-L/2}^{L/2} d\chi \int_{-\infty}^{\infty}\frac{dk}{2\pi}\int_{-\infty}^{\infty}\frac{dq}{2\pi}\bar{f}_k e^{i(k+q)\chi} \tag{9.84}$$

To begin with, we shall integrate over the space coordinates using the identity

$$\int_{-\infty}^{\infty} e^{i(k+q)\chi} d\chi = 2\pi\delta(k+q)$$

$$\langle E f_1\rangle = \lim_{L\to\infty}\frac{1}{L}\int_{-\infty}^{\infty}\frac{dk}{2\pi}\int_{-\infty}^{\infty} dq \bar{f}_k \bar{E}_q \delta(k+q) \tag{9.85}$$

Integrating the above equation over q, we get

$$\langle E f_1\rangle = \frac{1}{L_\infty}\int_{-\infty}^{\infty}\frac{dk}{2\pi}\bar{f}_k(t)\bar{E_{-k}}(t) \tag{9.86}$$

Note $\lim_{L\to\infty} 1/L = 1/L_\infty$ is the notation for the limit.

The linear solution for $\bar{f}_k$ is given by

$$\bar{f}_k(v,t) = \frac{e\bar{E}_k}{m_e(p+ikv)}\frac{\partial f_0}{\partial v} \tag{9.87}$$

The time evolution of $f_0(v,t)$ in terms of $\bar{E}_k$ is given by

$$\frac{\partial f_0}{\partial t} = \left(\frac{e}{m_e}\right)^2 \frac{\partial}{\partial v}\frac{1}{L_\infty}\int_{-\infty}^{\infty}\frac{dk}{2\pi}\frac{\bar{E}_k\bar{E_{-k}}}{(p+ikv)}\frac{\partial f_0}{\partial v} \tag{9.88}$$

The above equation looks like a diffusion equation, wherein the spatial coordinate is replaced by the velocity. In other words,

$$\frac{\partial f_0}{\partial t} = \frac{\partial}{\partial v} D \frac{\partial}{\partial v} f_0 \tag{9.89}$$

where

$$D = \frac{e^2}{m_e^2 L_\infty}\int_{-\infty}^{\infty}\frac{dk}{2\pi}\frac{\bar{E}_k\bar{E_{-k}}}{p+ikv} \tag{9.90}$$

Separating the integral of the above equation into its principal part and the pole term leads to

$$\lim_{b\to 0}\frac{1}{a+ib} = \frac{\mathcal{P}}{a} - i\pi\delta(a)$$

Equation (9.90) can be split into two parts, one being the principal part, while the second is the resonant term. Letting $p = \gamma - i\omega$, we have

$$\frac{\partial f_0}{\partial t} = \frac{\partial}{\partial v}\frac{e^2}{m_e^2 L_\infty}\int_{-\infty}^{\infty}\frac{dk}{2\pi}\left[i\mathcal{P}\left(\frac{\bar{E}_k\bar{E_{-k}}}{\omega - kv}\right) + \pi\delta(\omega - kv)\bar{E}_k\bar{E_{-k}}\right]\frac{\partial f_0}{\partial v} \tag{9.91}$$

© 2015 by Taylor & Francis Group, LLC

An interesting observation is that the two terms contribute differently to the wave-particle interaction. The principal part does not involve the particle with velocity given by $v = \omega/k$, it is nonresonant and has a very weak effect.

In weak turbulence theory, the basic conservation laws are due to the kinetic equation. However, even for the multicomponent environment of weak turbulence, both momentum and energy are conserved. This is a very useful concept and we shall very briefly outline the conservation of these two quantities. It is easy to check that the velocity distribution function is invariant as

$$\frac{d}{dt}\int dv f_{0j} = \int dv \frac{\partial f_{0j}}{\partial t} = 0 \tag{9.92}$$

However, for the momentum conservation, distribution function requires a sum over wave momentum and particle species. For electrostatic waves, this is straightforward as the waves carry no momentum and the rate of change of particle momentum density is given by

$$\begin{aligned}
\frac{d}{dt}\sum_j P_j(t) &= \sum_j \bar{n_j} m_j \int dv \mathbf{v} \frac{\partial f_{0j}}{\partial t} \\
&= -\sum_j \bar{n_j} m_j \int dv D_j \frac{\partial f_{0j}}{\partial v} \\
&= -\sum_\alpha \sum_j \frac{\bar{n_j} q_j^2}{m_j L_\infty} \int_{-\infty}^{\infty} \frac{dk}{2\pi} \int dv \frac{\bar{E_k^\alpha} E_{-k}^{\bar{\alpha}}}{(p^\alpha + ikv)} \frac{\partial f_0}{\partial v}
\end{aligned} \tag{9.93}$$

The middle result is derived by integrating by parts and the last expression from equation (9.90). The notation $\sum_\alpha$ is over the spectrum of unstable waves, with $p^\alpha = \gamma_k^\alpha - i\omega_k^\alpha$ and $\gamma_k^\alpha > 0$. From equation (9.88), one can show that the dispersion relation for the electrostatic waves is given by

$$\sum_j \frac{i\omega_{pj}^2}{k} \int \frac{\frac{\partial f_{0j}}{\partial v}}{p + ikv} dv = 1 \tag{9.94}$$

Equation (9.93) can be written as

$$\frac{d}{dt}\sum_j P_j(t) = 2i \sum_\alpha \int_{-\infty}^{\infty} \frac{dk}{2\pi} \mathcal{W}_k^\alpha k \tag{9.95}$$

where $\mathcal{W}_k^\alpha = \epsilon_0 |E_k^\alpha|^2 / 2L_\infty$ is the energy density of the wave as $E_{-k}^\alpha = E_k^\alpha$. As k is odd, and $\mathcal{W}_k^\alpha$ is even in k, the integral vanishes so that the momentum density is conserved. Finally, let us outline the conservation of energy density:

$$\frac{d}{dt}\sum_j (1/2)\bar{n}_j m_j \int dv v^2 f_j = -\sum_j \bar{n}_j m_j \int dv v D_j \frac{\partial f_{0j}}{\partial v}$$

© 2015 by Taylor & Francis Group, LLC

$$= 2\sum_{\alpha}\int \frac{dk}{2\pi}\mathcal{W}_k^{\alpha}\sum_j \frac{\omega_{pj}^2}{k^2}\int dv \frac{(ikv)ik\dfrac{\partial f_{0j}}{\partial v}}{p^{\alpha}+ikv}$$

Letting $ikv \to (ikv + p^{\alpha}) - p^{\alpha}$ in the numerator and using equation (9.94), we have

$$\frac{d}{dt}\sum_j (1/2)\bar{n}_j m_j \int dv v^2 f_j = -2\sum_{\alpha}\int \frac{dk}{2\pi}\mathcal{W}_k^{\alpha}\gamma_k^{\alpha} \tag{9.96}$$

Noting that the energy density varies with time as

$$\frac{\partial \mathcal{W}_k^{\alpha}}{\partial t} = 2\gamma_k^{\alpha}\mathcal{W}_k^{\alpha} \tag{9.97}$$

Equation (9.96) transforms to

$$\frac{d}{dt}\Big[\sum_j (1/2)\bar{n}_j m_j \int dv v^2 f_j + \sum_{\alpha}\int \frac{dk}{2\pi}\mathcal{W}_k^{\alpha}\Big] = 0 \tag{9.98}$$

The above expression clearly demonstrates that the sum of the particle energy and the wave energy is conserved. In the case of electromagnetic waves in the presence of a static magnetic field, the analysis is rather difficult and technical and we skip the analysis here.

9.9 Nonlinear Waves in Plasmas

In the previous section, we discussed the weakly nonlinear effects arising in weak turbulence. However, nonlinear effects play a dominant role in different types of waves. In the strongly nonlinear limit, shock waves may occur, when the wave energy density and the plasma energy densities are comparable. They are sometimes described by solitary waves, in the lower limit of amplitudes. A weakly nonlinear theory can sometimes lead to Landau damping due to trapping of particles. However, when the trapped particles begin to bounce on very short time scales, relative to the Landau damping rate, they will severely modify the energy balance between particles and waves, resulting in a change in the damping rate. For large amplitudes, the contribution due to damping ultimately ceases, so that the distribution function is modified in the vicinity of the phase velocity to support stationary waves. An important nonlinear effect that plays an important role is the pondermotive force. We shall briefly discuss the importance of this force before discussing the full nonlinear effects in plasmas.

© 2015 by Taylor & Francis Group, LLC

9.9.1 Pondermotive Force

For waves traveling in a plasma, particularly when the plasma or the wave amplitude is not homogeneous, the wave amplitude itself can effectively modify the plasma density profile through the pondermotive force. One of the consequences of this pondermotive force is that a localized wave can effectively expel plasma from the vicinity of the wave amplitude maximum, and this may tend to make the wave even more localized, leading to a trough in the plasma.

In an inhomogeneous, high frequency field, the motion of an electron may be divided into two parts which describe the high frequency oscillation about an effective center of oscillation and the relatively slow motion of this oscillation center. If we take the electric field to be given by $E(x,t) = E - 0(x)\cos t$, and imagine that the amplitude is increasing in the positive $x$direction, then as the electron moves into the stronger field, it will be accelerated more strongly back toward the oscillation center. However, as it moves toward the weaker field as x goes negative, it receives a weaker restoring force. On the average then, it will experience a slow drift toward the weaker field as if under the influence of a steady or slowly varying force, while at the same time experiencing the rapid oscillation at the high frequency.

Consider the equation of motion of a charged particle in an inhomogeneous electric field:

$$m\ddot{\mathbf{r}} = qE_0(\mathbf{r})\cos\omega t \tag{9.99}$$

The motion will be split into two (one a slow and the other a fast motion) so that $r = r_0 + r_1$, where $\langle r\rangle = r_0$ is the average over the fast time scale, or over the period $T = 2\pi/\omega/$. r_0 describes the oscillation center, while r_1 the rapidly oscillation motion, which is determined by

$$m\ddot{\mathbf{r_1}} = qE_0\cos\omega t \tag{9.100}$$

The solution of the above equation is trivial and is given by

$$r_1 = -(qE_0/m\omega^2)\cos\omega t \tag{9.101}$$

The slow variation can be expanded about r_0 so that the equation of motion reduces to

$$m(\ddot{\mathbf{r_0}} + \ddot{\mathbf{r_1}}) = q[\mathbf{E_0} + (\mathbf{r_1}\cdot\nabla)\mathbf{E_0}]\cos\omega t \tag{9.102}$$

Averaging the above equation over a period leads to

$$m\ddot{\mathbf{r_0}} = q\langle\mathbf{r_1}\cos\omega t\rangle\nabla\mathbf{E_0} \tag{9.103}$$

Using equation (9.101) for $\mathbf{r_1}$, the average can be simplified to yield

$$\langle\mathbf{r_1}\cos\omega t\rangle = -q\mathbf{E_0}/2m\omega^2$$

Equation (9.103) becomes

$$m\ddot{\mathbf{r_0}} = -\frac{q^2}{2m\omega^2}\mathbf{E_0}\cdot\nabla\mathbf{E_0} = -\frac{q^2}{4m\omega^2}\nabla(E_0^2) \tag{9.104}$$

© 2015 by Taylor & Francis Group, LLC

The above equation is the result of a pondermotive force, with its associated pondermotive potential given by

$$F_p = -\frac{q^2}{4m\omega^2}\nabla(E_0^2) = -\nabla\psi_p \tag{9.105}$$

$$\psi_p = \frac{q}{4m\omega^2}E_0^2 \tag{9.106}$$

Inclusion of inhomogeneous magnetic fields results in a drift of the guiding center in addition to the motion from the pondermotive force. However, the pondermotive force is unchanged.

9.9.2 Ion-Acoustic Solitons

To begin with, we shall define the following variables (normalized) as

$$n = n_i/n_0;\ \ u = v_i/c_s;\ \ \chi = \zeta/\lambda_D;\ \ \tau = c_s t/\lambda_D;\ \ \Phi = e\phi/\kappa T_e$$

The ion fluid equations can be written as

$$\frac{\partial n}{\partial \tau} + \frac{\partial}{\partial x}(nu) = 0 \tag{9.107}$$

$$\frac{\partial u}{\partial \tau} + u\frac{\partial u}{\partial x} = -\frac{\partial \Phi}{\partial x} \tag{9.108}$$

The normalized electron density is $n_e = exp(\Phi)$ so that the Poisson's equation becomes

$$\frac{\partial^2 \Phi}{\partial x^2} = e^{\Phi} - n \tag{9.109}$$

To start with, we shall linearize the equations to see the behavior of the solutions. Setting $\partial/\partial\tau \to -i\omega$, $\partial/\partial x \to ik$, $n = 1 + \bar{n}$, and $n_e = 1 + \bar{n_e}$ leads to

$$-i\omega\bar{n} + iku = 0 \Rightarrow u = \omega k\bar{n}$$

$$-i\omega u = -ik\Phi \Rightarrow \Phi = \frac{\omega}{k}u = \frac{\omega^2}{k^2}\bar{n}$$

$$k^2\Phi = \bar{n} - \Phi \Rightarrow k^2 = \frac{\omega^2}{1-\omega^2} \approx \omega^2(1 + \omega^2 + \omega^4 + \cdots)$$

As $\omega^2 \ll 1$ (ω is normalized to ω_{pi}) and $\omega/k \approx 1$ (normalized), we have

$$kx - \omega\tau \approx \omega x + (1/2)\omega^3 x - \omega\tau = \omega(x-\tau) + (1/2)\omega^3 x$$

At this stage, we shall introduce new variables $\eta = \omega(x-\tau)$, and $\xi = \omega^3 x$ and a small parameter $\epsilon = \omega^2$ so that

$$kx - \omega\tau \sim \epsilon^{1/2}(x-\tau) + (1/2)\epsilon^{3/2}x = \eta + (1/2)\xi \tag{9.110}$$

The change in variables results in

© 2015 by Taylor & Francis Group, LLC

$$\frac{\partial}{\partial x} = \frac{\partial \eta}{\partial x}\frac{\partial}{\partial \eta} + \frac{\partial \xi}{\partial x}\frac{\partial}{\partial \xi} = \epsilon^{1/2}\frac{\partial}{\partial \eta} + \epsilon^{3/2}\frac{\partial}{\partial \xi}$$

$$\frac{\partial}{\partial \tau} = \frac{\partial \eta}{\partial \tau}\frac{\partial}{\partial \eta} + \frac{\partial \xi}{\partial \tau}\frac{\partial}{\partial \xi} = -\epsilon^{1/2}\frac{\partial}{\partial \eta}$$

Expanding upto the second order in ϵ and using the basic equations as

$$n = 1 + \epsilon n^{(1)} + \epsilon^2 n^{(2)} + \cdots$$

$$u = 1 + \epsilon u^{(1)} + \epsilon^2 u^{(2)} + \cdots$$

$$\Phi = 1 + \epsilon \Phi^{(1)} + \epsilon^2 \Phi^{(2)} + \cdots$$

The first-order equations reduce to

$$\frac{\partial n^{(1)}}{\partial \eta} = \frac{\partial u^{(1)}}{\partial \eta} = \frac{\partial \Phi^{(1)}}{\partial \eta} \tag{9.111}$$

The solution of the above equation leads to $n^{(1)} = u^{(1)} = \Phi^{(1)} = \phi$. The second-order equations can be written as

$$\frac{\partial n^{(2)}}{\partial \eta} - \frac{\partial u^{(2)}}{\partial \eta} = \frac{\partial u^{(1)}}{\partial \xi} + \frac{\partial}{\partial \eta}(n^{(1)}u^{(1)}) = \phi_\xi + (\phi^2)_\eta$$

$$\frac{\partial u^{(2)}}{\partial \eta} - \frac{\partial \Phi^{(2)}}{\partial \eta} = \frac{\partial u^{(1)}}{\partial \xi} + u^{(1)}\frac{\partial u^{(1)}}{\partial \eta} = \phi_\xi + \phi\phi_\eta$$

$$\Phi^{(2)} - n^{(2)} = \frac{\partial^2 \Phi^{(1)}}{\partial \eta^2} - \frac{(\Phi^{(1)})^2}{2} = \phi_{\eta\eta} - \frac{1}{2}\phi^2$$

Differentiating the last equation with respect to η, adding all the equations, and eliminating second-order quantities results in

$$\phi_\xi + \phi\phi_\eta + \frac{1}{2}\phi_{\eta\eta\eta} = 0 \tag{9.112}$$

The above equation is the famous KdV equation with $a = 1$ and $b = 1/2$. The KdV equation is more than an ordinary solitary-wave equation since it has been proved that individual solitary waves survive collisions, and hence, these are called solitons to indicate their particle-like behavior. The KdV equation and other soliton equations may be solved by the method of inverse scattering which is a method of solving the nonlinear partial differential equations for arbitrary initial conditions by linear methods. The solution via inverse scattering for the KdV equation is discussed by Davidson[117], but for a more complete discussion of this method and its application to other soliton equations, Ablowitz and Segur[99] and Lamb[138] are useful references. Through this formalism, it has been possible to show that the solitary waves are indeed normal modes and that arbitrary perturbations invariably relax to these solitary waves and the remainder radiates away. Analytic N-soliton solutions are also known and there are an infinite number of conservation laws associated with the KdV equation and other soliton equations.

© 2015 by Taylor & Francis Group, LLC

9.9.3 Nonlinear Schrödinger Equation

A nonlinear wave where the wave amplitude is large enough to modify the dispersion relation by changing the parameters of the medium is governed by the nonlinear Schrodinger equation. A similar equation which admits soliton solutions using the method of inverse scattering has been named the Zakharaov−Shabat equation. We shall briefly mention both these equations and discuss the method of obtaining the solution of these equations.

Assume a plane wave of the form

$$Ae^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} + A^* e^{-(\mathbf{k}\cdot\mathbf{r}-\omega t)} \tag{9.113}$$

and a nonlinear dispersion relation (reasonably in general form)

$$D(\omega, \mathbf{k}, |A|^2) = 0 \tag{9.114}$$

Here it important to realize that the nonlinearity has to do with the amplitude rather than the phase of the wave. This kind of nonlinearity is quite typical of the pondermotive force. We also assume that the medium is homogeneous and isotropic in the absence of the wave, so that D and A are scalar quantities. We assume the nonlinear effects will result in a slow modulation of the wave amplitude about an average amplitude A_0, so we expand ω and k, which represent the time and space differential operators, about the fast variation values, so that equation (9.114) becomes a differential equation for the slowly varying amplitude:

$$Di\left[\omega + \mathbf{i}\frac{\partial}{\partial t}, \mathbf{k} - \mathbf{i}\nabla, |A(\mathbf{r}, t)|^2\right] A = 0 \tag{9.115}$$

Let us assume that the differential operators are small ($\omega \gg \partial/\partial t, k \gg \nabla$) and $|A^2| - |A_0|^2 \ll |A_0|^2$, so that

$$\left[\mathbf{i}\frac{\partial D}{\partial \omega}\frac{\partial}{\partial t} - \mathbf{i}\frac{\partial D}{\partial \mathbf{k}}\cdot\nabla - \frac{1}{2}\frac{\partial^2 D}{\partial \mathbf{k}\partial \mathbf{k}} : \nabla^2 + (|A|^2 - |A_0|^2\frac{\partial D}{\partial |A_0|^2} + \cdots\right] A = 0 \tag{9.116}$$

Dividing the above equation by $\partial D/\partial\omega$, we get

$$i\left(\frac{\partial}{\partial t} + v_g \cdot \nabla\right) A - P : \nabla^2 A + Q(|A|^2 - |A_0|^2)A = 0 \tag{9.117}$$

where the group velocity is given by $v_g = -(\partial D/\partial \mathbf{k})/(\partial D/\partial\omega) = \partial\omega/\partial\mathbf{k}$ and

$$P = \frac{1}{2}\frac{\partial v_g}{\partial \mathbf{k}} = \frac{1}{2}\frac{\partial^2 \omega}{\partial \mathbf{k}\partial \mathbf{k}}$$

$$Q = \frac{\partial D}{\partial |A_0|^2}\left(\frac{\partial D}{\partial \omega}\right)^{-1} = -\frac{\partial \omega}{\partial |A_0|^2}$$

Here Q represents the nonlinear frequency shift due to the finite amplitude of

© 2015 by Taylor & Francis Group, LLC

the wave. Let us introduce a moving frame. Moving with the group velocity as $\xi = \mathbf{r} - v_g t$, we have the nonlinear Schrodinger equation as

$$\mathbf{i}\frac{\partial A}{\partial t} + P : \nabla_\xi^2 A + Q(|A|^2 - |A_0|^2)A = 0 \tag{9.118}$$

The above equation is the three-dimensional nonlinear Schrodinger equation, whereas for the one-dimensional case, the equation reduces to

$$\mathbf{i}\frac{\partial A}{\partial t} + P\frac{\partial^2 A}{\partial \xi^2} + Q(|A|^2 - |A_0|^2)A = 0 \tag{9.119}$$

The case when $A_0 = 0$ is called the Zakharov−Shabat equation in the literature. The nature of the solutions for both the equations are completely different.

9.9.4 Zakharaov−Shabat Equation

Consider the nonlinear evolution equation given by

$$\mathbf{i}u_t + Pu_{\xi\xi} + Q|u|^2 u = 0 \tag{9.120}$$

The subscript denotes differentiation with the respective variable. Assume a wave packet solution of the form

$$u = e^{ik(\xi - v_p t)}\Phi[K(\xi - v_g' t)] \tag{9.121}$$

Note that $\xi - v_g' t = x - (v_g + v_g')t$. v_g' represents the excess velocity of the wave packet over the linear group velocity. Inserting equation (9.121) into equation (9.120) and separating the real and imaginary parts, we have

$$kv_p\Phi + P(k^2\Phi'' - k^2\Phi) + Q\Phi^3 = 0 \tag{9.122}$$

$$-iK\Phi'(v_g' - 2Pk) = 0 \tag{9.123}$$

Here Φ' denotes the derivative of Φ with respect to its argument. The wavenumber can be written as $k = v_g'/2P$. Simplifying equation (9.122), we get

$$\frac{1}{2}\frac{d}{d\Phi}(\Phi')^2 = a\Phi - 2b\Phi^3 \tag{9.124}$$

where

$$a = (k^2 - kv_p)/PK^2$$

$$b = Q/2PK^2$$

Separation of variables and simplifying the integral leads to

$$K(\xi - v_g' t) = \int \frac{d\phi}{\Phi\sqrt{a - b\Phi^2}} = -\frac{1}{\sqrt{a}}\sec h^{-1}\left(\frac{\sqrt{b}\Phi}{\sqrt{a}}\right)$$

© 2015 by Taylor & Francis Group, LLC

with the solution given by

$$\begin{aligned}\Phi(\xi,t) &= \sqrt{a/b}\,\mathrm{sec}\,h[\sqrt{a}K(\xi - v_g' t)] \\ &= \Phi_m \,\mathrm{sec}\,h[K(\xi - v_g' t)] \end{aligned} \tag{9.125}$$

Here, $a = 1$ and $b = 1/\Phi_m^2$, resulting in $K = \Phi_m\sqrt{Q/2P}$ and $v_p = (v_g'/2) - \Phi_m^2 QP/v_g'$. The normal form of the Zakharov−Shabat equation has $P = 1$ and $Q = 2$, so that $K = \Phi_m$, $k = v_g'/2$ and $v_p = (v_g'/2) - 2\Phi_m^2/v_g'$.

We have now come to the end of the discussion on waves in fluids and plasmas. The next chapter will deal with instabilities in fluids and plasmas. Our discussion in the last three sections of this chapter, which dealt with nonhomogeneous plasmas, quasi-linear theory and nonlinear waves is very close to the treatment given in Swanson[170]. However, that book is very technical and one needs a good mathematical background to have a good grasp of it.

9.10 Exercises

1. In a plasma whose temperature $T \approx 1.6 \times 10^6 K$ and $n \approx 6.0\mathrm{cm}^{-3}$, calculate the Debye length.

2. Calculate the charge in a plasma whose plasma frequency is given by $8KHz$, mass of the electron $m_e = 9.1 \times 10^{-31} Kg$, and $n = 6\mathrm{cm}^{-3}$.

3. Calculate the cyclic gyro frequency in a plasma whose charge is $q = 1.6 \times 10^{-19}$Coulumbs, $m_e 9.1 \times 10^{-31}$Kg for different values of the magnetic field $B = 0.5, 10, 50,$ Gauss.

4. For a collision frequency ν_c proportional to $1/v^3$, where v is the electron speed, derive the relation between ν_c and the kinetic energy.

5. For an unmagnetized plasma, assume that $\chi = -n_e^2/m\omega^2$. Show that the dielectric constant ϵ is given by $1 - \omega_{pe}^2/\omega^2$. Discuss briefly the limiting case of $\epsilon \to 1$.

6. For electromagnetic waves in a nonmagnetic plasma, write the dispersion relation in terms of the dielectric constant. Discuss the nature of the phase and group velocities.

7. The upper hybrid frequency is defined by $\omega_{UH}^2 = \omega_{pe}^2 + \omega_{Be}^2$. Derive the cut-off for the x mode by solving for the frequency ω.

8. If the dispersion relation for warm plasma waves is

$$\omega^2 = \omega_{pe}^2 + (3/2)k^2 v_{th}^2$$

© 2015 by Taylor & Francis Group, LLC

what can one say about the group velocity? What happens in the case of cold plasma?

9. The relation for the polarized (right and left) waves in terms of μ is given by

$$\mu^2 = 1 - \frac{\omega_{pe}^2/\omega^2}{1 \pm \omega_{Be}/\omega}$$

Discuss the nature of μ for low frequency ($\omega \to 0$) and ($\omega \to \infty$).

10. Show that the equation

$$\frac{d^2y}{dx^2} + \beta^2(x - x_0)y = 0$$

with the transformation $z = -\beta^{2/3}(x - x_0)$ reduces to an equation of the form $y'' - \zeta y = 0$.

© 2015 by Taylor & Francis Group, LLC

Chapter 10

Fluid and Plasma Instabilities

10.1 Introduction 451
10.2 Stability of Parallel Shear Flows 452
10.2.1 Squire's Theorem 454
10.2.2 Rayleigh's Inflexion-Point Theorem 455
10.2.3 Fjortoft's Theorem 455
10.2.4 Howard's Semi-Circle Theorem 456
10.3 Taylor−Goldstein Equation 456
10.4 Orr−Sommerfeld Equation 461
10.5 Rayleigh–Taylor (RT) Instability 464
10.5.1 Application 470
10.6 Kelvin−Helmholtz (KH) Instability 470
10.6.1 Application 474
10.7 Parametric Instability 475
10.7.1 Application 477
10.8 Two-Stream Instability 478
10.9 Interchange (Flute) Instability 482
10.10 Sausage Instability 485
10.11 Kink Instability 486
10.12 Ballooning Instability 487
10.13 Exercises 490

10.1 Introduction

In Chapters 1 through 9, we have been discussing waves and oscillations in different media, both homogeneous and inhomogeneous. Also, we have discussed waves in fluids and plasmas, considering the different types of external forces such as gravity, pressure gradient, rotation, electric and magnetic fields. We have also briefly discussed two fluid description in plasmas. Needless to say, the complete discussion was mostly concentrated towards the linear theory. However, wherever possible, we have mentioned some nonlinear studies also. An important question that arises in dealing with the waves and oscillations is that of stability. We have derived the dispersion relation for the different physical systems and discussed the nature of the modes of the system. The most important aspect that we have not so far addressed is the behavior of the perturbations (both small and finite). Are these perturbations stable or unstable? We shall address these questions in this chapter. However, we shall restrict ourselves only to linear stability theory and skip the nonlinear stabil-

© 2015 by Taylor & Francis Group, LLC

ity aspects. That discussion is highly technical and is beyond the scope of this book.

Analysis on the stability of flows were formulated as early as the work of Reynolds in the 19th century. The method of normal modes to study the oscillations and the stability of a dynamical system was reasonably understood. For example, one sought a solution of Newton's or Lagrange's equations of motion. Under idealized situations, explicit, analytical solutions were obtained. The need to study the behavior of the perturbations of these solutions was greatly felt. The process was to linearize the equations of motion by neglecting the products of the perturbations. The basic assumption was that the perturbation of each quantity could be resolved into independent components, i.e., modes with varying time t like $e^{\omega t}$, where ω was a constant, which in general could be complex. If the real part of the constant ω was found to be positive for any mode of the system, then this mode was deemed unstable as the initial perturbations, as functions of time, would grow exponentially until they cease to remain small. This approach which is often referred to as the normal mode approach was utilized by Stokes, Kelvin, and Rayleigh during the late nineteenth century, for fluid dynamic problems.

Very few solutions to nonlinear equations of motion are known. Even more complicated is the analysis pertaining to their instability. Thus, one is forced to study the stability of a few classes of simpler laminar flows, which have planar, axial, or spherical symmetry. One of the main reasons for instability in a system is due to the fact that a disturbance in a fluid which is in equilibrium is caused by external forces, such as inertia and viscous stresses in the fluid. Some of the external forces that we may wish to study are the buoyancy in a stratified system (variable density), surface tension, and MHD forces (electric and magnetic fields). One may also have to study the contribution due to the Coriolis forces as an external force when the system is rotating and in which the fluid is moving. In the absence of an external force or viscosity, a fluid will move according to the equilibrium between its inertia and internal stresses of pressure. A minor perturbation may upset the equilibrium so that the disturbance is either amplified or damped.

In this chapter, we shall start the discussion on the stability of parallel shear flows, which are purely two-dimensional and inviscid. We shall generalize this discussion to include effects arising due to external forces such as viscosity, buoyancy, and magnetic field.

10.2 Stability of Parallel Shear Flows

Let us assume that the basic flow for a parallel two-dimensional flow be is the form

© 2015 by Taylor & Francis Group, LLC

$$\hat{\mathbf{U}} = \hat{U}(z)\mathbf{i} \qquad (z_1 < z < z_2) \tag{10.1}$$

where $\mathbf{i}$ denotes the unit vector in the x direction. It is assumed that the basic shear flow $\hat{U}(z)$ is an arbitrary function of z. The flow is assumed to be bounded by the two planes $z = z_1$ and $z = z_2$, which may be either rigid or free. For a rigid boundary, the boundary condition is that the normal component of the velocity must vanish. However, for a free boundary, the pressure must be constant. In principle, one of the boundaries may be at infinity as in the case of boundary layers or both may be at infinity as in the case of shear layers, jets, etc.

It is important to nondimensionalize the governing equations of motion. For this, we shall introduce a characteristic length L and a characteristic velocity V, associated with the basic flow. This choice may not be unique. A reasonable assumption would be

$$V = max_{z_1<z<z_2}|\hat{U}(z)| \tag{10.2}$$

For flows in a channel, one can choose $L = (1/2)(z_2 - z_1)$. Define the nondimensional quantities as

$$t = \hat{t}V/L, \quad \mathbf{x} = \hat{\mathbf{x}}/L, \quad \mathbf{u} = \hat{\mathbf{u}}/V, \quad p = \hat{p}/\rho V^2, \quad \mathbf{U} = \hat{\mathbf{U}}/V = U(z)\mathbf{i} \tag{10.3}$$

Here, ρ is the density of the fluid. The Euler's equation of motion can be written as

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u} = -\nabla p \tag{10.4}$$

$$\nabla\cdot\mathbf{u} = 0 \tag{10.5}$$

It is easy to verify that the basic flow, which is an arbitrary function of the vertical coordinate z, automatically satisfies the boundary conditions. Also the equation (10.4) is automatically satisfied provided $\nabla p = 0$. This means that the pressure has to be constant. In order to study the stability of this flow, we shall assume

$$\mathbf{u}(\mathbf{x},t) = U(z)\mathbf{i} + \mathbf{u}'(\mathbf{x},t) \tag{10.6}$$

$$p(\mathbf{x},t) = constant + p'(\mathbf{x},t) \tag{10.7}$$

where $\mathbf{u}'$ and p' are the disturbance in the velocity and pressure, respectively. By a simple algebra, the linearized equations of motion can be written as

$$\Big(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\Big)\mathbf{u}' + w'\frac{dU}{dz}\mathbf{i} = -\nabla p' \tag{10.8}$$

$$\nabla\cdot\mathbf{u}' = 0 \tag{10.9}$$

An important point to be noted at this stage regards studying the above

© 2015 by Taylor & Francis Group, LLC

system as an initial value problem. The caution is that this is rather complicated while the other approach of normal mode analysis of the system of equations (10.8 and 10.9) is reasonably easier. However, the normal modes of the linearized inviscid stability problem are far from being complete. They do provide some role with regard to the initial value problem approach. In the normal mode approach, we shall assume the solutions of the system as

$$\mathbf{u}'(\mathbf{x},t) = \hat{\mathbf{u}}(z)\exp[i(kx+my-\omega t)] \tag{10.10}$$

$$p'(\mathbf{x},t) = \hat{p}(z)\exp[i(kx+my-\omega t)] \tag{10.11}$$

We have to take the real parts of these expressions to obtain the physical quantities. For the solutions to remain bounded, the wavenumbers k and m must be real. The frequency can, in general, be complex, $\omega = \omega_r + i\omega_i$. The above expressions thus represent waves which travel in the direction $(k, m, 0)$ with a phase speed given by $\omega/(k^2+m^2)^{1/2}$, and may grow or decay in time like $\exp(\omega_i t)$. This wave is designated as being stable if $\omega_i \leq 0$ and unstable if $\omega_i > 0$. It is called neutrally stable if $\omega_i = 0$. If we introduce the symbol $c = \omega/k$, then c can be interpreted as the phase speed of the disturbance. If we also introduce $D = d/dz$, then the basic equations can be written as a system of ordinary differential equations:

$$ik(U-c)\hat{u} + U'\hat{w} = -ik\hat{p} \tag{10.12}$$

$$ik(U-c)\hat{v} = -im\hat{p} \tag{10.13}$$

$$ik(U-c)\hat{w} = -D\hat{p} \tag{10.14}$$

$$i(k\hat{u}+m\hat{v}) + D\hat{w} = 0 \tag{10.15}$$

The boundary conditions for the rigid boundaries are given by

$$\hat{w} = 0, \quad at \quad z = z_1, \quad \text{and} \quad z_2 \tag{10.16}$$

Applying the boundary condition into the system of ordinary differential equations and simplifying leads to the dispersion eigenvalue relation of the form

$$\mathcal{F}(k,m,c) = 0 \tag{10.17}$$

For a given basic flow $U(z)$ and a given vector wavenumber (k, m, c), the above relation determines the possible values of c. The allowed values of c, in general, are finite in number and are thus not complete. In addition to the discrete spectrum, there is a continuous part which arises from the singularity in the equations when the term $(U - c) = 0$. For the sake of completeness, we shall state here, without proof some of the theorems which were derived by several authors.

10.2.1 Squire's Theorem

To each unstable three-dimensional disturbance, there corresponds a more unstable two - dimensional one.

© 2015 by Taylor & Francis Group, LLC

An interesting alternative is the induction of a stream function $\psi'(x, y, t)$ such that the two components of the velocity are given by

$$u' = \partial\psi'/\partial z, \quad \text{and} \quad w' = -\partial\psi/\partial x \tag{10.18}$$

In addition, we assume that

$$\psi'(x, z, t) = \phi(z)e^{i(x-ct)} \tag{10.19}$$

then

$$\hat{u} = \phi', \quad \textit{and} \quad \hat{w} = -ik\phi \tag{10.20}$$

Equation (10.12) reduces to

$$\hat{p} = U'\phi - (U - c)\phi' \tag{10.21}$$

Substituting the above expression into equation (10.14), we have

$$(U - c)(\phi'' - k^2\phi) - U''\phi = 0 \tag{10.22}$$

The above equation is often referred to in the literature as the Rayleigh stability equation. The corresponding boundary conditions are given by

$$k\phi = 0, \quad \textit{at} \quad z = z_1, \quad \textit{and} \quad z_2 \tag{10.23}$$

An interesting property is that the above equation is symmetric, in the sense that it does not change by changing k to $-k$. The adjoint of the Rayleigh stability equation can be written as

$$(D^2 - k^2)(U - c)\phi^* - U''\phi^* \tag{10.24}$$

where ϕ^* satisfies the same boundary conditions. The self-adjoint form of the equation can be written as

$$D[(U - c)^2 D\phi^*] - k^2(U - c)^2\phi^* = 0 \tag{10.25}$$

We shall state without proof the consequences of the above equation.

10.2.2 Rayleigh's Inflexion-Point Theorem

A necessary and sufficient condition for instability is that the basic velocity profile should have an inflexion point.

10.2.3 Fjortoft's Theorem

A necessary condition for instability is that $U''(U - U_s) < 0$ *somewhere in the field of flow, where* z_s *is a point at which* $U'' = 0$, *and* $U_s = U(z_s)$

It is interesting to note that the Rayleigh stability equation does posseses a neutral stable eigensolution with $k_s > 0$. Assume that $K(z) = -U''/(U - U_s)$

© 2015 by Taylor & Francis Group, LLC

is a regular function at z_s. This implies that $U_s'' = 0$. Assume $\lambda = -k^2$. Assuming that $c = c_s = U_s$, the Rayleigh stability equation can be rewritten as

$$\phi'' + \lambda + K(z)\phi = 0 \tag{10.26}$$

The above equation with the corresponding boundary conditions (10.23) can be written in a standard Sturm−Liouville problem which admits an infinite sequence of eigenvalues with the limit point tending to ∞. The minimum of this spectrum is given by

$$\lambda_s = min \int_{z_1}^{z_2} (f^2 - Kf^2)dz / \int_{z_1}^{z_2} f^2 dz \tag{10.27}$$

where the minimum is taken for the functions f which satisfy the boundary conditions which have square integrable derivatives. From the inequality

$$(z_2 - z_1)^2 \int_{z_1}^{z_2} f'^2 dz \geq \pi^2 \int_{z_1}^{z_2} f^2 dz \tag{10.28}$$

we can show that

$$K(z) > \pi^2/(z_2 - z_1)^2 \tag{10.29}$$

everywhere. Thus, instability is possible only when $0 < k < k_s$ and we have stability ($c_i = 0$) for $k \geq k_s$. Another consequence is Howard's semi-circle theorem which is as follows.

10.2.4 Howard's Semi-Circle Theorem

For unstable waves, c must lie in the semicircle given by

$$[c_r - (1/2)(U_{max} + U_{min})]^2 + c_i^2 \leq [(1/2)(U_{max} - U_{min})]^2 \quad (c_i > 0)$$

10.3 Taylor−Goldstein Equation

A straightforward generalization to the above discussion is to include the effect of gravity (uniform) so that the discussion would be carried out for a stratified fluid in the vertical direction z. Let us begin the discussion of the Taylor−Goldstein with an application to the Atmosphere of the Earth. A model having a smooth wind and density profiles for which a solution of the Taylor−Goldstein equation is possible will be described here. This discussion will be for a simple three-layer model in which the velocity has a piecewise linear profile. The approach is similar to the normal mode approach as discussed in the previous section. The procedure is identical and so we shall skip

© 2015 by Taylor & Francis Group, LLC

the details for the sake of brevity. Small perturbations obey the wave equation given by (Gossard and Hooke[128])

$$\frac{d^2W(z)}{dz^2} + \left[\frac{N^2}{(U-c)^2} - \frac{U''}{(U-c)} - k^2 - \frac{2\Gamma U'}{(U-c)} - \Gamma^2\right] W(z) = 0 \quad (10.30)$$

The expression $W(z) = [\rho(z)/\rho_0]^{1/2}w$, where w is the vertical component of the velocity perturbation, ρ is the density (unperturbed), while ρ_0 is the density at some specific level. The square of the Brunt–Vaisala frequency is given by $N^2 = gd(\ln\theta)/dz$, where θ is the potential temperature. The ambient shear and its second derivative with respect to the height z are given by U and U''. c is the phase velocity of the disturbance, while Γ is given by

$$\Gamma = (g/2c_s^2)(1 - N^2c_s^2/g^2) \quad (10.31)$$

where c_s is the speed of sound. Whenever c_s is large, the terms containing Γ are negligible as compared with the first term in the coefficient for reasonable values of U'. In this case, the wave equation reduces to

$$\frac{d^2W(z)}{dz^2} + \left[\frac{N^2}{(U-c)^2} - \frac{U''}{(U-c)} - k^2\right] W(z) = 0 \quad (10.32)$$

The above equation is the famous equation studied by Taylor and Goldstein in 1931 and subsequently by many other authors. This approximation is also equivalent to the Boussinesq approximation wherein the variation of the density is neglected except in the gravity term. It is important to realize that whenever $U = c$, the two terms in the wave equation become infinite, which leads to critical levels (the levels at which the basic shear and the phase velocity coincide). One option to circumvent this situation is to assume that N^2 and U'' go to zero at these levels, rapidly enough to overcome the zeros at these levels. Another option is to let $W(z)$ go to zero when $z \to z_c$.

We shall assume the stratification in terms of a potential temperature $\theta(z)$ instead of the density $\rho(z)$. For the fluid to be stable statically, it is imperative that the potential temperature θ increase with height, whereas in the case of the density $\rho(z)$, it should decrease. The typical profiles for the potential temperature and ambient shear are as follows:

$$\theta = \theta_0 \exp\left[N_\infty^2 \frac{\triangle H}{g}\left(\frac{z}{\triangle H} - \tan h\frac{z}{\triangle H}\right)\right] \quad (10.33)$$

$$U = U_\infty \tan h\frac{z}{\triangle H} \quad (10.34)$$

where N_∞^2 and U_∞ are the Brunt–Vaisala frequency and ambient shear, respectively as $z \to \infty$.

The choice of the smooth profile model is advantageous as it can be compared with a piecewise smooth velocity profile (linear, for example). However,

© 2015 by Taylor & Francis Group, LLC

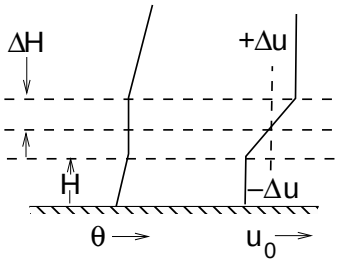


FIGURE 10.1: Three-layered model due to Gossard[22].

such profiles may sometimes represent actual atmospheric conditions during the onset of intense mixing associated with wind discontinuity. The mixing of the momentum would tend to erase the shear as well as the gradient of potential temperature in the eventual steady state, but if the shear is maintained by some process external to the model, it is more easily amenable to stability analysis. We also assume that

$$N^2 = N_\infty^2 \tan h^2 \frac{z}{\triangle H} \tag{10.35}$$

$$U'' = -2\frac{U_\infty}{\triangle H^2} \sec h^2 \frac{z}{\triangle H} \tan h \frac{z}{\triangle H} \tag{10.36}$$

both of which vanish at $z = 0$.

The wave equation may be applied to a two-dimensional disturbance propagating in the x direction. Thus, $c = \omega/k$, where ω is the frequency and k the wavenumber in the direction. Solutions for which the frequency (and hence the phase speed) is complex will represent unstable perturbations. In this section, we are interested more in the real solutions of c as such solutions define the stability boundary. In addition, the profiles are symmetric so that we can assume, without loss of generality, $c = 0$. The wave equation now reduces to

$$\frac{d^2W(z)}{dz^2} + \frac{1}{(\triangle H)^2}\left[\left(\frac{N_\infty}{\beta_0}\right)^2 - \alpha^2 + 2\sec h^2 \frac{z}{\triangle H}\right] W(z) = 0 \tag{10.37}$$

where $\beta_0 = U_\infty/\triangle H$ and $\alpha = k\triangle H$. The solution of the above equation satisfying the boundary conditions $W(\pm\infty) = 0$ is given by

$$W(z) = A \sec h \frac{z}{\triangle H} \tag{10.38}$$

$$\left(\frac{N_\infty}{\beta_0}\right)^2 = \alpha^2 - 1 \tag{10.39}$$

In what follows, we shall compare the stability boundary, which has smooth profiles with continuous derivatives, with the stability boundary of a corresponding three-layer, which is a piecewise linear model due to Gossard[22].

© 2015 by Taylor & Francis Group, LLC

The wind and potential temperature profiles for Gossard's three-layer model are presented in Figure 10.1 . This presents the general eigensolution for the three-layer model, piecewise linear (almost) but continuous in shear and potential temperature with a rigid boundary. The relation between ω and α in terms of H, $\triangle H$, N_1, N_3, and β is given by

$$\left[\left(1+\frac{\alpha_1\alpha_3}{\alpha^2}\right)\tan h2\alpha+\frac{\alpha_1+\alpha_3}{\alpha}\right]\left(\frac{\omega}{\beta}\right)^2+\left[\left(\frac{\alpha_1-\alpha_3}{\alpha}\right)\tan h2\alpha\right]\frac{\omega}{\beta}$$

$$+(2\alpha-\alpha\alpha_1-\alpha\alpha_3)-(1+\alpha^2-\alpha_1-\alpha_3+\alpha_1\alpha_3)\tan h2\alpha=0 \qquad (10.40)$$

where

$$\alpha_1=\gamma_1\triangle H\cot h\gamma_1 H$$

$$\alpha_3=\gamma_3\triangle H$$

$$\gamma_1=k\left[1-\left(\frac{N_1/\beta}{\omega/\beta+\alpha}\right)^2\right]^{1/2}$$

$$\gamma_3=k\left[1-\left(\frac{N_3/\beta}{\omega/\beta+\alpha}\right)^2\right]^{1/2}$$

The subscripts 1 and 3 apply to the lower and upper layers, respectively. N_1 and N_3 are assumed to be constant in the lower and upper layers, so that the profiles for the potential temperature θ are exponential and thus not quite linear above and below the shear layer. It is assumed that $N_2 = 0$ in the middle layer. For the special case when $N_1 = N_3$, $H = \infty$, and $\omega/\beta = c$, this becomes the three−layer analog of the smooth model. For this case equation (10.40) reduces to

$$[(\alpha_\infty-1)^2+\alpha^2]\tan h2\alpha+2\alpha(\alpha_\infty-1)=0 \qquad (10.41)$$

where $\alpha_\infty=\alpha^2-(N_\infty/\beta)^2$ and the subscript ∞ designates the value in both layers 1 and 3. The stability curve plotted from the above equation is shown in Figure 10.2. The general shape is the same as that of the solid curve plotted from equation (10.41). Both curves pass through $(N_\infty/\beta_0)^2=-1$ at $\alpha=0$.

Another interesting solution of the wave equation is given by

$$W(z)=A\cos h^2(z/\triangle H) \qquad (10.42)$$

The drawback of this solution is that it does not satisfy the boundary conditions and so we shall not consider it. The solution (equation 10.41) defines a curve shown as the solid line in Figure 10.2 . This depicts a locus of stable solution in the $\alpha,(N_\infty/\beta_0)^2$ plane, which implies the existence of a continuous region of unstable solutions. Thus, equation (10.41) defines the stability boundary. The unstable region lies above this curve.

When the model has a homogeneous density distribution, it becomes a

© 2015 by Taylor & Francis Group, LLC

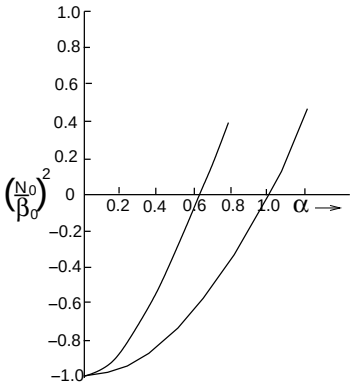


FIGURE 10.2: Neutral curves:solid for the present model; dashed for the three-layer model Gossard's [22].

special case as studied in the Section 10.2.3 on the Rayleigh condition. When $N_\infty = 0$ and $\alpha_\infty = \alpha$, the above equation reduces to the Rayleigh condition for stability which is given by

$$\frac{\tan h 2\alpha}{1 + \tan h 2\alpha} - \alpha(1 - \alpha) > 0 \tag{10.43}$$

which yields $\alpha > 0.64$. Thus, the stability boundary for the three−layer model passes through 0.64 at $N_\infty/\beta = 0$ as seen in the Figure 10.2.

It is interesting to note that for the potential temperature profile given by

$$\theta = \theta_0 \left[N_0^2 \frac{\triangle H}{g} \tan h \left(\frac{z}{\triangle H} \right) \right] \tag{10.44}$$

where N_0 is the Brunt–Vaisala frequency at $z = 0$, the Richardson number given by the ratio of the Brunt–Vaisala frequency and the square of the gradient of the shear do not vanish at the critical level.

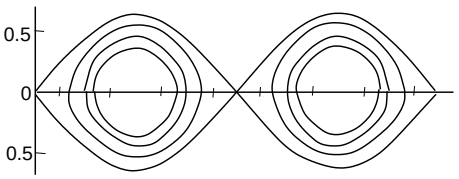


FIGURE 10.3: Streamline pattern of this model.

© 2015 by Taylor & Francis Group, LLC

The eigensolutions and stability boundaries are given by

$$W(z) = A\left[\sec h\left(\frac{z}{\triangle H}\right)\right]\left[\tan h\left(\frac{z}{\triangle H}\right)\right]^{1-\epsilon} \tag{10.45}$$

$$\left(\frac{N_0}{\beta}\right)^2 = \epsilon(1-\epsilon) \tag{10.46}$$

The above equation has a maximum at $(N_0/\beta)^2 = (1/4)$. This is in good agreement with the Miles−Howard theorem which states that a sufficient condition for stable fluid flow is that the local Richardson number $[(N/\beta)^2]$ be greater than 1/4 everywhere in the fluid. The stability boundary curves shown in the Figure 10.3, however, not exhibit such a limit. This is because $(N/\beta)^2 = 0$ at $z = 0$, which is a contradiction to the Miles−Howard theorem.

At $N_0 = N_\infty = 0$, corresponding to a homogeneous fluid or an adiabatic atmosphere, all three models give a value of $\alpha = 1$ as the stability boundary. The piecewise model due to Gossard[22]. however, yields a value of $\alpha = 0.64$, as mentioned above. The reason for such a difference is due to the basic shear profiles. The stream lines for the present model resemble that of a cat's eyes as shown in Figure 10.3. The cat's eye pattern occurs whenever $W(z_c) \neq 0$. This means that all terms of the Taylor−Goldstein equation remain finite at the critical level $z = z_c$. Thus the existence of the cat's eye depends critically on the distribution of the Brunt–Vaisala frequency wherein it becomes zero at the critical level. Mathematically this is a special case. However, there are several observational records for the cat's eye pattern in the atmosphere.

10.4 Orr−Sommerfeld Equation

In this section, we shall introduce the effect of viscosity and for simplicity deal with only two-dimensional incompressible, unsteady flow and obtain an equation whose solutions can be used to predict the wavelengths of the disturbances which will amplify in a shear layer. The basic equations of motion are the Navier-Stokes equations given by

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{10.47}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \frac{1}{\rho}X - \frac{1}{\rho}\frac{\partial p}{\partial x} + \nu\left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right] \tag{10.48}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = \frac{1}{\rho}Y - \frac{1}{\rho}\frac{\partial p}{\partial y} + \nu\left[\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right] \tag{10.49}$$

where X and Y are some general external forces perturbing the velocity and the pressure terms as $u = U + u'$, $v = v'$, $p = P + p'$. It is easy to check

© 2015 by Taylor & Francis Group, LLC

that the mean flow automatically satisfies the Navier-Stokes equations. Thus, one can subtract the mean flow equation to get the equations in terms of the perturbations. We shall neglect body-force terms and the quadratic terms in the perturbations. Also we assume that

$$\frac{\partial U}{\partial x} \ll \frac{\partial U}{\partial y} \tag{10.50}$$

The perturbed equations of motion can be written as

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} = 0 \tag{10.51}$$

$$\frac{\partial u'}{\partial t} + U\frac{\partial u'}{\partial x} + v'\frac{\partial U}{\partial y} + \frac{1}{\rho}\frac{\partial p'}{\partial x} = \nu\nabla^2 u' \tag{10.52}$$

$$\frac{\partial v'}{\partial t} + U\frac{\partial v'}{\partial x} + \frac{1}{\rho}\frac{\partial p'}{\partial y} = \nu\nabla^2 v' \tag{10.53}$$

Assuming a periodic disturbance and also the fact that they can be expressed in terms of a Fourier series, the solution for each term can be obtained and summed up (this is possible due to the linear assumption) and is valid for small amplitude disturbances.

For two-dimensional flows, we can define a stream function as follows:

$$\psi(x, y, t) = \Phi(y)e^{i(kx-\omega t)} \tag{10.54}$$

where the wavenumber $k = 2\pi/l$. Here l is the wavelength of the disturbance.

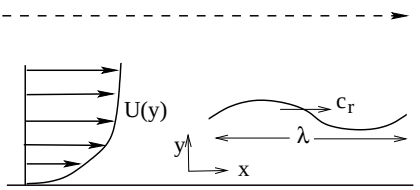


FIGURE 10.4: A simple sketch of the shear layer.

The frequency $\omega = \omega_r + i\omega_i$ can be split as a real and complex part. ω_i will be the amplification factor. We can as usual define $c = \omega/k = c_r + ic_i$, the complex phase speed. c_r may be defined as the propagation velocity in the x direction, while c_i is the degree of damping. If c_i is negative, then it is amplification. The perturbed velocity components can be written as

$$u' = \frac{\partial \Phi}{\partial y} = \Phi'(y)e^{i(kx-\omega t)} \tag{10.55}$$

$$v' = \frac{\partial \Phi}{\partial x} = -ik\Phi'(y)e^{i(kx-\omega t)} \tag{10.56}$$

© 2015 by Taylor & Francis Group, LLC

Substituting into the basic equations (perturbed) and simplifying, we have the following equation:

$$(U-c)(\Phi''-k^2\Phi)-U''\Phi=-\frac{1}{kR_e}(\Phi''''-2k^2\Phi''+k^4\Phi) \tag{10.57}$$

The above equation is the famous Orr−Sommerfeld equation. All the lengths are divided by a characteristic length δ of the mean flow (for example, a channel flow) or by the boundary layer or shear layer thickness. The velocities are divided by a quantity U_m, the maximum velocity of the given flow. The Reynolds number is defined as

$$R_e=\frac{U_m\delta}{\nu} \tag{10.58}$$

In the above equation, the terms on the left-hand side are the inertial terms which are of the second order, while the terms on the right-hand side are the viscous terms which are of the fourth order. A simple sketch of the shear layer is shown in Figure 10.4. The boundary conditions are given by

$$y = 0, \quad \text{at} \quad u'=v'=0; \quad \text{implies} \quad \Phi=0, \quad \Phi'=0 \tag{10.59}$$

$$y = \infty, \quad \text{at} \quad u'=v'=0; \quad \text{implies} \quad \Phi=0, \quad \Phi'=0 \tag{10.60}$$

If we try to solve the frictionless stability equation first, then we regain the stability results of Rayleigh and the theorems on the velocity profile and the inflection point. The result of such a solution is presented in Figure 10.5.

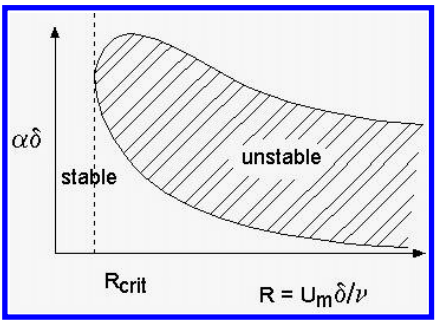


FIGURE 10.5: The solution for the frictionless case.

The vertical coordinate is the ratio of the shear layer thickness to the wavelength of the disturbance, while the horizontal axis is the Reynolds number based on the maximum (free stream) velocity and the shear layer thickness. The curve in Figure 10.5 shows the boundaries between regions of stability (where disturbances damp out) and regions of instability (where disturbances amplify). It should be very obvious to us that it is called the thumb curve, as it looks typically like a thumb. For a critical Reynolds number, all disturbances dies out so that the shear layer is stable for any small disturbance.

© 2015 by Taylor & Francis Group, LLC

However, for a Reynolds number greater than this critical number, there are instabilities which will amplify. For a larger Reynolds number, most of the disturbances lead to turbulence. The inviscid instability shown in the figure requires that the velocity profile have an inflection point as is necessary per Rayleigh's theorem. However, under some specified conditions, there can be an instability, even without an inflection point in the velocity. This is usually termed the viscous instability. In order to model such a system. one has to include viscous terms on the right-hand side of the Navier-Stokes equations. Figure 10.6 shows the region of viscous instability, where oscillations develop in spite of the absence of an inflexion point in the velocity profile. This region gets narrower to a point as the Reynolds number tends to ∞ as is evident from the figure.

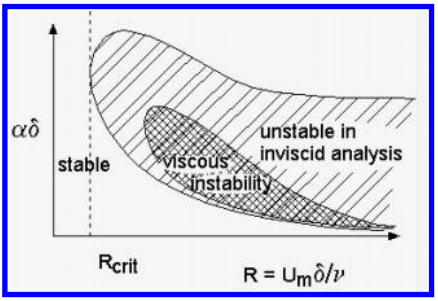


FIGURE 10.6: The solution with the inclusion of friction.

10.5 Rayleigh–Taylor (RT) Instability

We will start a discussion on RT instability initially for hydrodynamics and then move on to the MHD. For example, let us consider a situation wherein a heavy block is kept on a lighter block and achieves equilibrium. If we were to ask , whether such a configuration is stable, the answer will be negative. In hydrodynamics, a small rippling motion would disturb the equilibrium and make it unstable. The ripples are unstable because they effectively interchange volume elements of heavy fluid with equivalent volume elements of lighter fluid. Each volume element of interchanged heavy fluid originally had its center of mass a distance $\triangle$ above the interface while each volume element of interchanged light fluid originally had its center of mass a distance $\triangle$ below the interface. The potential energy of a mass m at a height h is given by mgh, where g is the gravitational field. With the respective changes in the potential energy of the lighter and heavier fluid, with simplifications, we get

© 2015 by Taylor & Francis Group, LLC

$$\delta W_h = -2\rho_h V \triangle g, \qquad \delta W_l = +2\rho_l V \triangle g \tag{10.61}$$

V in the above expression is the volume of the interchanged fluid elements and ρ_h and ρ_l are the densities (mass) of the heavy and lighter fluid, respectively. The net change in the potential energy of the total system is

$$\delta W = -2(\rho_h - \rho_l) V \triangle g \tag{10.62}$$

which is less than zero. Thus, the system lowers its potential energy by forming ripples. This is similar to a ball rolling from the top of the hill.

A well-known example of this instability is that of the inverted glass of water. The heavy fluid in this case is the water while the air is the lighter fluid. This system will be stable when a piece of cardboard is located at the interface between the water and the air. However, with the removal of the cardboard, the system becomes unstable and the water starts to fall out. The cardboard prevents the ripple interchange from happening. The system remains stable when the ripples are prevented, for the atmospheric pressure is adequate to support the inverted water. The constraint on the system is the cardboard, with a boundary condition that prevents ripple formation. However, when the cardboard is removed, there is no longer any constraint against ripple formation. Thus, the ripples tend to grow into large amplitudes with the result that water falls. This is a typical case of an unstable equilibrium. The reader may look into the following books on topics related to Rayleigh – Taylor (RT) and Kelvin−Helmholtz (KH) instabilities, Chandrasekhar[113], Chen[114], Bellan[103], and Dendy[118].

Let us work on the above argument in hydrodynamics with the usual equations of motion, so that we can get a quantitative idea on the stability of such a system. Consider the stability of a heavier fluid such as water, supported by air which is lighter, with no constraint at the interface. Assume the vertical in the y direction so that gravity acts in the negative direction. Let us assume that $\rho_l \ll \rho_h$, so that the mass of the lighter fluid can be neglected. The water and air are assumed to be incompressible and have no variation in the density. The continuity equation is

$$\frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho + \rho \nabla \cdot \mathbf{v} = 0 \tag{10.63}$$

In the case of an incompressible fluid, it reduces to

$$\nabla \cdot \mathbf{v} = 0 \tag{10.64}$$

The continuity equation (linearized) in the water reduces to

$$\frac{\partial \rho_1}{\partial t} + \mathbf{v}_1 \cdot \nabla \rho_0 = 0 \tag{10.65}$$

while the linearized equation of motion in the water is

$$\rho_0 \frac{\partial \mathbf{v}_1}{\partial t} = -\nabla P_1 - \rho_1 g \hat{y} \tag{10.66}$$

© 2015 by Taylor & Francis Group, LLC

The unperturbed interface between air and water is defined by the line $y = 0$ while the top of the glass is at $y = h$. The boundary condition for the water at the top would be

$$v_y = 0 \quad at \quad y = h \tag{10.67}$$

Assuming perturbations of the form

$$\mathbf{v}_1 = \mathbf{v}_1(y)e^{\gamma t + i\mathbf{k}\cdot\mathbf{x}} \tag{10.68}$$

where $\mathbf{k}$ lies in the x-z plane and the positive γ implies instability, the incompressibility condition reduces to the following equation:

$$\frac{\partial v_{1y}}{\partial y} + i\mathbf{k}\cdot\mathbf{v}_{1\perp} = 0 \tag{10.69}$$

where $\perp$ means perpendicular to the y direction. The y and $\perp$ components of equation (10.69) become, respectively,

$$\gamma\rho_0 v_{1y} = -\frac{\partial P_1}{\partial y} - \rho_1 g \tag{10.70}$$

$$\gamma\rho_0\mathbf{v}_{1\perp} = -i\mathbf{k}p_1 \tag{10.71}$$

Applying the dot product of equation (10.71) with $i\mathbf{k}$ and using equation (10.69), by eliminating $\mathbf{k}\cdot\mathbf{v}_{1\perp}$, we obtain

$$\gamma\rho_0\frac{\partial v_{1y}}{\partial y} = k^2 P_1 \tag{10.72}$$

The perturbed density, as given in equation (10.65) is

$$\gamma\rho_1 = -v_{1y}\frac{\partial\rho_0}{\partial y} \tag{10.73}$$

Substituting the expressions for ρ_1 and P_1 into equation (10.69), we obtain the eigenvalue problem

$$\frac{\partial}{\partial y}\left[\gamma^2\rho_0\frac{\partial v_{1y}}{\partial y}\left[\gamma^2\rho_0 - g\frac{\partial\rho_0}{\partial y}\right]k^2 v_{1y}\right. \tag{10.74}$$

The interior and interface are solved separately from the above equation. For the interior, $\partial\rho_0/\partial y = 0$ and $\rho_0 = constant$, which means that the equation (10.74) reduces to

$$\frac{\partial^2 v_{1y}}{\partial y^2} = k^2 v_{1y} \tag{10.75}$$

with the solution satisfying the boundary condition as

$$v_{1y} = A\sin h[k(y-h)] \tag{10.76}$$

© 2015 by Taylor & Francis Group, LLC

To derive the condition at the interface, equation (10.74) should be integrated from $y = 0_-$ to $y = 0_+$ to obtain

$$\left[\gamma^2 \rho_0 \frac{\partial v_{1y}}{\partial y}\right]_{0_-}^{0_+} = -\left[g\rho_0 k^2 v_{1y}\right]_{0_-}^{0_+} \tag{10.77}$$

or

$$\gamma^2 \frac{\partial v_{1y}}{\partial y} = -gk^2 v_{1y} \tag{10.78}$$

where all quantities refer to the upper (water) side of the interface, since by assumption $\rho_0(y = 0_-) \approx 0$. Substituting equation (10.76) into equation (10.78) the dispersion relation can be derived as

$$\gamma^2 = (kg) \tan h[k_\perp h] \tag{10.79}$$

The above expression implies that the configuration is always unstable whenever $\gamma^2 > 0$. It is also important to realize that the short wavelengths are the most unstable ones. Taking into account other effects such as surface tension, the system can be made stable for a given range of wavelengths. A simple illustration of the Rayleigh – Taylor instability is sketched in Figure 10.7.

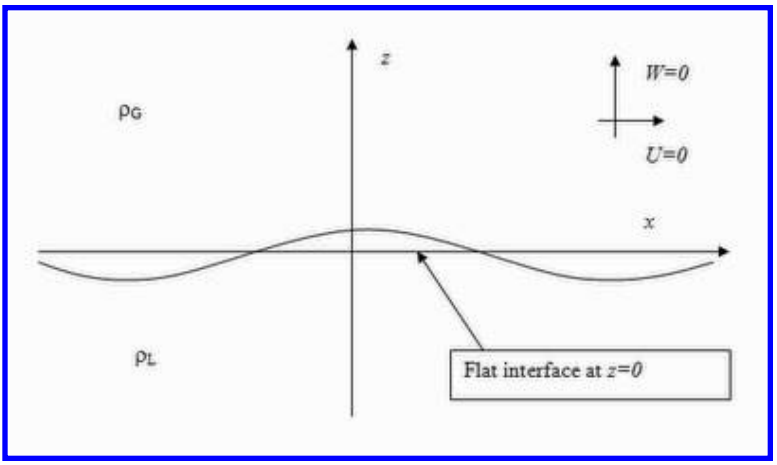


FIGURE 10.7: A simple sketch of the Rayleigh–Taylor configuration.

We turn our attention to RT instability in magneto hydrodynamic fluids (MHD). We shall replace water by a magneto fluid (a fluid which satisfies the MHD equations and atmospheric pressure, by a vertical magnetic field whose gradient balances the gravitational force, i.e., at each y, the upward force of $-\nabla b^2/2\mu_0$ supports the downward force of the weight of the plasma. In order for $-\nabla b^2$ to point upwards in the y direction, the magnetic field must depend on y such that its magnitude decreases with increasing y. Also it is required that $B_y = 0$ so that ∇B^2 is perpendicular to the magnetic field and the field

© 2015 by Taylor & Francis Group, LLC

can be considered as locally straight. Assuming the equilibrium magnetic field to be

$$\mathbf{B}_0 = B_{x0}(y)\hat{x} + B_{z0}(y)\hat{z} \tag{10.80}$$

the equilibrium unit vector may be written as

$$\hat{B_0} = \frac{B_{x0}(y)\hat{x} + B_{z0}(y)\hat{z}}{\sqrt{B_{x0}(y)^2 + B_{z0}(y)^2}} \tag{10.81}$$

For the special case when $B_{x0}(y)$ and $B_{z0}(y)$ are proportional to each other, the field line becomes independent of the y direction, while for the general case, $\hat{B_0}$ would depend on y, which rotates as a function of y. For the incompressible fluid, the linearized equations of motion can be written as

$$\rho_0 = -\nabla \bar{P}_1 + \frac{\mathbf{B}_0 \cdot \nabla \mathbf{B}_1 + \mathbf{B}_1 \cdot \mathbf{B}_0}{\mu_0} - \rho_1 g \hat{y} \tag{10.82}$$

where

$$\bar{P}_1 = P_1 + \frac{\mathbf{B}_0 \cdot \ \mathbf{B}_1}{\mu_0} \tag{10.83}$$

is the perturbation of the total (hydrodynamic and magnetic) pressure, i.e., $P + B^2/2\mu_0$. The equation (10.82) with its components is given by

$$\gamma \rho_0 v_{1y} = -\frac{\partial \bar{P}}{\partial y} + \frac{i(\mathbf{k} \cdot \mathbf{B}_0) B_{1y}}{\mu_0} - \rho_1 g \tag{10.84}$$

$$\gamma \rho_0 \mathbf{v}_{1\perp} = -i\mathbf{k}\bar{P} + \frac{1}{\mu_0}\left[i(\mathbf{k} \cdot \mathbf{B}_0)\mathbf{B}_{1\perp} + B_{1y}\frac{\partial \mathbf{B}_0}{\partial y} \right] \tag{10.85}$$

Performing dot product of equation (10.85) with $i\mathbf{k}$ and using equation (10.69) we obtain

$$-\gamma \rho_0 \frac{\partial v_{1y}}{\partial y} = k^2 \bar{P} + \frac{1}{\mu_0}\left[-(\mathbf{k} \cdot \mathbf{B}_0)\mathbf{k} \cdot \mathbf{B}_{1\perp} + iB_{1y}\frac{\partial (\mathbf{k} \cdot \mathbf{B}_0)}{\partial y} \right] \tag{10.86}$$

With the magnetic field being divergent free, i.e., $\nabla \cdot \mathbf{B}_1 = 0$, the perturbed perpendicular field can be written as

$$i\mathbf{k} \cdot \mathbf{B}_{1\perp} = -\frac{\partial B_{1y}}{\partial y} \tag{10.87}$$

so that equation (10.86) reduces to

$$k^2 \bar{P} = -\gamma \rho_0 \frac{\partial v_{1y}}{\partial y} - \frac{1}{\mu_0}\left[-i(\mathbf{k} \cdot \mathbf{B}_0)\frac{\partial B_{1y}}{\partial y} + iB_{1y}\frac{\partial (\mathbf{k} \cdot \mathbf{B}_0)}{\partial y} \right] \tag{10.88}$$

Eliminating $\bar{P}$ from equation (10.84) and substituting into the above equation leads to the following:

© 2015 by Taylor & Francis Group, LLC

$$\gamma\rho_0 v_{1y} = -\frac{1}{k^2}\frac{\partial}{\partial y}\left[-\gamma\rho_0\frac{\partial v_{1y}}{\partial y} - \frac{1}{\mu_0}\left[-(i\mathbf{k}\cdot\mathbf{B}_0)\frac{\partial B_{1y}}{\partial y} + B_{1y}\frac{\partial(i\mathbf{k}\cdot\mathbf{B}_0)}{\partial y}\right]\right] + \frac{i(\mathbf{k}\cdot\mathbf{B}_0)B_{1y}}{\mu_0} - \rho_1 g \tag{10.89}$$

Using Ohm's law, we obtain

$$\mathbf{E}_1 + \mathbf{v}\times\mathbf{B}_0 = 0 \tag{10.90}$$

Taking curl and incorporating Faraday's law, we get

$$\gamma\mathbf{B}_1 = \nabla\times[\mathbf{v}_1\times\mathbf{B}_0] \tag{10.91}$$

Taking dot product with $\hat{y}$ and using the vector identity

$$\nabla\cdot(\mathbf{F}\times\mathbf{G}) = \mathbf{G}\cdot\nabla\times\mathbf{F} - \mathbf{F}\cdot\nabla\times\mathbf{G} \tag{10.92}$$

we obtain

$$\gamma B_{1y} = \hat{y}\cdot\nabla\times[\mathbf{v}\times\mathbf{B}_0] = \nabla\cdot[(\mathbf{v}\times\mathbf{B}_0)\times\hat{y}] = \nabla\cdot[v_{1y}\mathbf{B}_0] = i\mathbf{k}\cdot\mathbf{B}_0 v_{1y} \tag{10.93}$$

Substituting into equation (10.89), using equation (10.91) and equation (10.73), and simplifying the terms gives

$$\frac{\partial}{\partial y}\left[\left[\gamma^2\rho_0 + \frac{1}{\mu_0}(\mathbf{k}\cdot\mathbf{B}_0)^2\right]\frac{\partial v_{1y}}{\partial y}\right] = k^2\left[\gamma^2\rho_0 - g\frac{\partial\rho_0}{\partial y} + \frac{(\mathbf{k}\cdot\mathbf{B}_0)^2}{\mu_0}v_{1y}\right] \tag{10.94}$$

When $\mathbf{k}\cdot\mathbf{B}_0 = 0$, the above equation reduces to the hydrodynamic case. On integration of the equation from $y = 0$ to $y = h$, we obtain

$$\left[\left(\gamma^2\rho_0 + \frac{1}{\mu_0}(\mathbf{k}\cdot\mathbf{B}_0)^2\right)v_{1y}\frac{\partial v_{1y}}{\partial y}\right]_0^h - \int_0^h\left[\gamma^2\rho_0 + \frac{1}{\mu_0}(\mathbf{k}\cdot\mathbf{B}_0)^2\right]\frac{\partial v_{1y}}{\partial y})^2 dy = k^2\int_0^h\left[\gamma^2\rho_0 - g\frac{\partial\rho_0}{\partial y} + \frac{(\mathbf{k}\cdot\mathbf{B}_0)^2}{\mu_0}\right]v_{1y}^2 dy \tag{10.95}$$

The first term on the left-hand side of the above equation is zero (on applying the boundary conditions), so the value for γ^2 turns out to be

$$\gamma^2 = \frac{\int_0^h dy\left[k^2 g\frac{\partial\rho_0}{\partial y}v_{1y}^2 - \frac{(\mathbf{k}\cdot\mathbf{B}_0)^2}{\mu_0}\left(k^2 v_{1y}^2 + \left(\frac{\partial v_{1y}}{\partial y}\right)^2\right)\right]}{\int_0^h dy\rho_0\left[k^2 v_{1y}^2 + \left(\frac{\partial v_{1y}}{\partial y}\right)^2\right]} \tag{10.96}$$

© 2015 by Taylor & Francis Group, LLC

For the case when $\mathbf{k} \cdot \mathbf{B}_0 = 0$, the density gradient is positive everywhere and yields $\gamma^2 > 0$ so that the instability sets in. If the density gradient is negative everywhere except at a region of finite thickness $\triangle y$, then the system will be unstable with respect to an interchange at one of the stratum. The velocity will be concentrated at this unstable strata and so the integrands will vanish everywhere except at the unstable stratum giving a growth rate $\gamma^2 \sim g \triangle y \rho_0^{-1} \partial \rho_0 / \partial y$, where $\partial \rho_0 / \partial y$ is the value in the unstable region.

In the literature, the hydromagnetic Rayleigh – Taylor instability is often referred to as the Kruskal−Schwarzschild instability, because for nomzero $\mathbf{k} \cdot \mathbf{B}_0$, opposes the effect of the destabilizing positive density gradient reducing the growth rate to $\gamma^2 \sim g \triangle y \rho_0^{-1} \partial \rho_0 / \partial y - (\mathbf{k} \cdot \mathbf{B}_0)^2 / \mu_0$. The important conclusion is that a sufficiently strong field will stabilize the system.

10.5.1 Application

The launch of the Hinode satellite led to the discovery of rising plumes, dark in chromospheric lines, that propagate from large (10 Mm) bubbles that form at the base of quiescent prominences. The plumes move through a height of approximately 10 Mm while developing highly turbulent profiles. The magnetic Rayleigh – Taylor instability was hypothesized to be the mechanism that drives these flows. Recently, Hillier et al.[91] using three-dimensional (3D) MHD simulations, investigated the nonlinear stability of the Kippenhahn−Schluter prominence model for the interchange mode of the magnetic Rayleigh – Taylor instability. Their model deals with the rise of a buoyant tube inside the quiescent prominence model, where the interchange of magnetic field lines becomes possible at the boundary between the buoyant tube and the prominence. Nonlinear interaction between plumes plays an important role for determining the plume dynamics. Using the results of ideal MHD simulations, they determined the initial parameters for the model and found that buoyancy affects the evolution of instability. They find that the 3D mode of the magnetic Rayleigh – Taylor instability grows, creating up flows aligned with the magnetic field of constant velocity (maximum found 7.3 km s^{-1}). The width of the upflows is dependent on the initial conditions, with a range of 0.5−4 Mm which propagates through heights of 3−6 Mm. Another application of RT stability has been studied by Ali et al.[92] in dense magneto plasmas.

10.6 Kelvin−Helmholtz (KH) Instability

In the presence of velocity shears at boundaries, we have low-frequency MHD waves being excited, which are called Kelvin-Helmholtz waves. The KH instabilities are different from RT instabilities. Unlike the Rayleigh – Taylor

© 2015 by Taylor & Francis Group, LLC

instability, which occurs due to density discontinuities, the (KH) instabilities have to do with velocity shears. A tangential discontinuity supports velocity shears and waves observed at such interfaces are called the KH waves. These waves are also termed as low frequency surface waves which can grow and become unstable under certain conditions. In this section, we describe the conditions for the onset of KH instability.

Let us assume a boundary across which there is a sheared flow. For simplicity, let the two fluids be ideal (infinite conductivity, $\sigma = \infty$),and incompressible, and the pressure isotropic. The equilibrium and perturbed quantities may be denoted by

$$\mathbf{v} = \mathbf{V}_0 + \delta\mathbf{V} \tag{10.97}$$

$$p = p_0 + \delta p \tag{10.98}$$

$$\mathbf{B} = \mathbf{B}_0 + \delta\mathbf{B} \tag{10.99}$$

Let the z axis be directed along the normal to the plane of the discontinuity, in a Cartesian system of geometry. For a tangential discontinuity $\mathbf{V}_0$ and $\mathbf{B}_0$ lying in the xy-plane, assume that the perturbation produces waves propagating in the xy plane with the condition that the waves may decay in strength away from the xy plane, in the z direction.

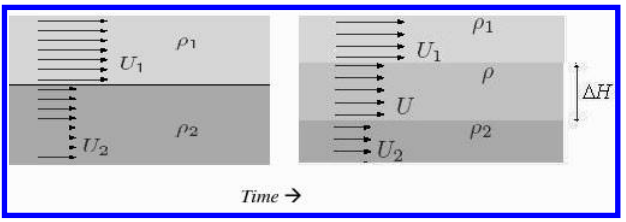


FIGURE 10.8: A simple sketch for the Kelvin−Helmholtz instability, from Gramer[23].

A simple sketch which describes the Kelvin−Helmholtz instability configuration is presented in Figure 10.8. If we assume that perturbations are of the form

$$\exp[i(k_x x + k_y y - \Omega t) - k_z z] \tag{10.100}$$

where Ω is the Doppler shifted frequency of the wave as measured by a stationary observer in the frame of the boundary, the decay length is given by the reciprocal of k_z. Thus, for $z < 0$, $k_z < 0$ and vice-versa. The ideal MHD equations are given by the following set of equations:

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} \tag{10.101}$$

$$\mathbf{E} = -\mathbf{V} \times \mathbf{B} \tag{10.102}$$

$$\rho_m \frac{d\mathbf{V}}{dt} = -\nabla p + \mathbf{J} \times \mathbf{B} \tag{10.103}$$

© 2015 by Taylor & Francis Group, LLC

To begin with, we assume that the pressure is scalar. The magnetic field and the current density are related to each other by the simple relation

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}. \tag{10.104}$$

We seek plane wave solution for equation (10.101). This results in

$$\begin{aligned} -i\Omega\delta\mathbf{B} &= (\mathbf{B}\cdot\nabla)\mathbf{V} - (\mathbf{V}\cdot\nabla)\mathbf{B} = (\mathbf{B}_0\cdot\nabla)\delta\mathbf{V} - (\mathbf{V}_0\cdot\nabla)\delta\mathbf{B} \\ &= i\mathbf{k}_t\cdot(B_0\delta\mathbf{V} - V_0\delta\mathbf{B}) \end{aligned} \tag{10.105}$$

The Doppler-shifted frequency is given by $\Omega = \omega + \mathbf{k}\cdot\mathbf{V}_0$, so that the above equation can be simplified as

$$\omega\delta\mathbf{B} = -(\mathbf{B}_0\cdot\mathbf{k}_t)\delta\mathbf{V} \tag{10.106}$$

For an incompressible fluid, the divergence-free condition has to be satisfied. This condition $\nabla\cdot\mathbf{V} = 0$ yields

$$\kappa_\pm\cdot\delta\mathbf{V} = 0 \tag{10.107}$$

where $\kappa_\pm = (k_x\hat{x} + K_y\hat{y}) \pm k_z\hat{z}$ is the complex wavenumber and κ_+ is for $z > 0$ and κ_- is for $z < 0$.

The equation of motion is simplified as

$$\rho_m d\mathbf{V}/dt = -\omega\rho_m\delta\mathbf{V} \tag{10.108}$$

$$\nabla p = \kappa_\pm\delta p \tag{10.109}$$

and

$$(\nabla\times\mathbf{B})\times\mathbf{B} = (\mathbf{B}\cdot\nabla)\mathbf{B} - \nabla B^2/2 \tag{10.110}$$

The first-order equations can be simplified so that the last term reduces to

$$(\mathbf{B}_0\cdot\nabla)\mathbf{B}_1 + \nabla(B_0^2/2 + \mathbf{B}_0\cdot\delta\mathbf{B}) = (\mathbf{B}_0\cdot\kappa_\pm)\delta\mathbf{B} - i\kappa_\pm(\mathbf{B}_0\cdot\delta\mathbf{B}) \tag{10.111}$$

Using the above expressions, equation (10.103) becomes

$$\omega\rho_m\delta\mathbf{V} - \kappa_\pm\delta p = -\frac{(\mathbf{B}_0\cdot\mathbf{k}_t)\delta\mathbf{B}}{\mu_0} + \frac{(\mathbf{B}_0\cdot\delta\mathbf{B})\kappa_\pm}{\mu_0} \tag{10.112}$$

The expression for the total pressure is $p^\star = p + B2/2\mu_0$ and to the first order

$$\delta p^\star = \delta p + \frac{\mathbf{B}_0\cdot\delta\mathbf{B}}{\mu_0} \tag{10.113}$$

Combining equation (10.112) and equation (10.106) results in

$$\kappa_\pm\delta p^\star = \frac{\rho_m}{\omega}\left[\omega^2 - \frac{(\mathbf{B}_0\cdot\mathbf{k}_t)^2}{\mu_0\rho_m}\right]\delta\mathbf{V} \tag{10.114}$$

© 2015 by Taylor & Francis Group, LLC

Performing the scalar product of the above equation with $\kappa_\pm$ and noting that $\kappa_\pm \cdot \delta\mathbf{V} = 0$, we get the following condition:

$$\kappa_\pm^2 \delta p^\star = 0 \tag{10.115}$$

The corresponding solution is given by

$$\delta p^\star = 0 \tag{10.116}$$

which yields the intermediate Alfven mode, i.e., $\omega/k = \pm V_A$, where the Alfven speed $V_A = B_0\lambda/\sqrt{\mu_0\rho_m}$ or

$$\kappa_\pm^2 = 0 \tag{10.117}$$

The second option is related to surface waves and yields

$$k_t^2 = k_z^2 \tag{10.118}$$

Equation (10.118) is a condition for the decay length of the perturbation away from the surface of discontinuity, i.e., $k_z^{-1} = k_t^{-1}$. Scalar multiplication of equation (10.114) with $\hat{\mathbf{z}}$ yields

$$\delta p_t \kappa_\pm \cdot \hat{\mathbf{z}} = \rho_m \left[\omega^2 - \frac{(\mathbf{B}_0 \cdot \mathbf{k}_t)^2}{\mu_0\rho_m}\right] \delta\mathbf{V} \cdot \hat{\mathbf{z}} \tag{10.119}$$

An important observation is to note that

$$\kappa_\pm \cdot \hat{\mathbf{z}} = \pm i k_z \tag{10.120}$$

and

$$\begin{aligned} \delta\mathbf{V} \cdot \hat{\mathbf{z}} &= \frac{d\delta z}{dt} = \frac{\partial \delta z}{\partial t} + (\mathbf{V}_0 \cdot \nabla)\delta z \\ &= i(\Omega - k\mathbf{V}_0)\delta z = i\omega\delta z \end{aligned} \tag{10.121}$$

Thus, equation (10.119) becomes

$$\pm k_z \delta p^\star = [\omega^2 - (\mathbf{V}_A \cdot \mathbf{k})^2]\rho_m \delta z \tag{10.122}$$

The boundary conditions are that the total pressure across a tangential discontinuity is continuous, i.e., $[p^\star] = 0$. Also, the displacement of the two fluids in the z direction must be continuous in order to avoid separation or inter penetration of the fluids, i.e., $[\delta z] = 0$. equation (10.122) has the form $\delta p^\star = A\delta z$. Thus, $[\delta p^\star] = [A\delta z] = [A][\delta z[= 0$, which implies that $[A] = 0$, which explicitly written will look as

$$\rho_1[\omega_1^2 - (\mathbf{V}_{A1} \cdot \mathbf{k})^2] + \rho_2[\omega_2^2 - (\mathbf{V}_{A2} \cdot \mathbf{k})^2] = 0 \tag{10.123}$$

The frequency shifted and measured by an observer is given by $\Omega = \omega_1 + \mathbf{V}_1 \cdot$

© 2015 by Taylor & Francis Group, LLC

$\mathbf{k} = \omega_2 + \mathbf{V}_2 \cdot \mathbf{k}$ and is the same in both sides. Solving for the expression Ω yields

$$\Omega = \frac{\rho_1 \mathbf{V}_1 \cdot \mathbf{k} + \rho_2 \mathbf{V}_2 \cdot \mathbf{k}}{(\rho_1 + \rho_2)} \pm \frac{1}{(\rho_1 + \rho_2)} \times \left[(\rho_1 + \rho_2) \Big[\rho_1 (\mathbf{V}_{A1} \cdot \mathbf{k})^2 + \rho_2 (\mathbf{V}_{A2} \cdot \mathbf{k})^2 \Big] - \rho_1 \rho_2 (\triangle \mathbf{V} \cdot \mathbf{k})^2 \right]^{1/2} \tag{10.124}$$

where $\triangle \mathbf{V} = \mathbf{V}_1 - \mathbf{V}_2$ is the basic shear, while the subscripts 0 and m have been omitted from the expression for velocity and density.

The dispersion relation for the Kelvin−Helmholtz waves is given by the expression (equation 10.124). These waves are unstable when Ω is imaginary. Simplifying the equation (10.124) shows that when

$$\rho_1 \rho_2 (\triangle \mathbf{V} \cdot \mathbf{k})^2 > \frac{1}{\mu_0} (\rho_1 + \rho_2) [(\mathbf{B}_1 \cdot \mathbf{k})^2 + (\mathbf{B}_2 \cdot \mathbf{k})^2] \tag{10.125}$$

the imaginary part of the relation (10.124) > 0.

For instability, a shear threshold is essential. This threshold is required because the tension in the magnetic field will resist any force acting on it to stretch it. This implies that the growth of the wave will occur only if the velocity shear overcomes the magnetic tension. If $\mathbf{B}_1$ and $\mathbf{B}_2$ are perpendicular to $\mathbf{k}$, then the right hand side of equation (10.124) vanishes and $(\triangle \mathbf{V} \cdot \mathbf{k})^2 > 0$ implies that the boundary is unstable to arbitrary small shear across the boundary. It is interesting to note that the growth rate depends on the relative directions of $\mathbf{B}$, $\mathbf{k}$, and $\triangle \mathbf{V}$. If α is the angle between $\triangle \mathbf{V}$ and $\mathbf{k}$ and θ_1, and θ_2 are the angles between $\mathbf{k}$ and $\mathbf{B}_1$ and $\mathbf{B}_2$, then the instability criterion can be written as

$$\triangle U^2 \cos^2 \alpha > \frac{1}{\mu_0} \left(\frac{1}{\rho_1} + \frac{1}{\rho_2} \right) (B_1^2 \cos^2 \theta_1 + B_2^2 \cos^2 \theta_2) \tag{10.126}$$

10.6.1 Application

One of the important applications of the K-H instability is in coronal streamers and may also be found in solar wind, where the plasma streams relative to the surrounding plasma which is relatively stationary. Flows and instabilities play a major role in the dynamics of magnetized plasmas including the solar corona, magnetospheric and heliospheric boundaries, cometary tails, and astrophysical jets. The nonlinear effects, multiscale and microphysical interactions inherent to the flow-driven instabilities, are believed to play a role, e.g., in plasma entry across a discontinuity, generation of turbulence, and enhanced drag. However, in order to clarify the efficiency of macroscopic instabilities in

© 2015 by Taylor & Francis Group, LLC

these processes, we lack proper knowledge of their overall morphological features. Foullon et al.[93] reported the first observations of the temporally and spatially resolved evolution of the magnetic Kelvin−Helmholtz instability in the solar corona. Unprecedented high-resolution imaging observations of vortices developing at the surface of a fast coronal mass ejecta were taken from the new Solar Dynamics Observatory, validating theories of the nonlinear dynamics involved. The new findings are a cornerstone for developing a unifying theory on flow-driven instabilities in rarefied magnetized plasmas, which is important for understanding the fundamental processes in key regions of the Sun−Earth system. Ofman and Thomson[94] have reported SDO (Solar Dynamics Observatory) observation of the Kelvin−Helmholtz instability in the solar corona.

10.7 Parametric Instability

One of the most thoroughly investigated nonlinear wave-wave interactions is the parametric instability. The theory, which is basically linear about an oscillating equilibrium. A standard discussion on the parametric instability is that of two oscillators, M_1 and M_2, which are coupled on a bar resting on a pivot. The pivot has the freedom to move forward and backward with a frequency ω_0, whereas the natural frequencies of the oscillators are ω_1 and ω_2, respectively. The pivot does not encounter any resistance as long as the masses M_1 and M_2 are not in motion. If P does not move, while M_2 is set to motion, it will induce movement to M_1, so long as the natural frequency of M_1 is different from ω_2, and the amplitude is small. If both P and M_2 are set in motion, then the displacement of M_1 as a function of time will be

$$\cos(\omega_2 t)\cos(\omega_0 t) = (1/2)\cos[(\omega_2 + \omega_0)t] + (1/2)\cos[(\omega_2 - \omega_0)t] \quad (10.127)$$

In the event of ω_1 equaling either $\omega_2 + \omega_0$ or $\omega_2 - \omega_0$, M_1 will get excited resonantly which will result in the growth of the amplitude. With M_1 oscillating, M_2 will gain energy because one of the beat frequencies of ω_1 with ω_0 is nothing but ω_2. Thus, if one of the oscillators starts moving, then the other will get excited which will result in the system becoming unstable. Depending on the energy coming from P, the oscillation amplitude is unaffected by M_1 and M_2, and the instability can be treated in the linear regime. In a plasma or MHD, P, M_1, and M_2 may be different types of waves.

Let us work out a more quantitative analysis pertaining to parametric instabilities by considering the equations of motion of two harmonic oscillators, x_1 and x_2 as follows:

$$\frac{d^2x_1}{dt^2} + \omega_1^2 x_1 = 0 \quad (10.128)$$

where ω_1 is the resonant frequency. If oscillator is driven by a time-dependent

© 2015 by Taylor & Francis Group, LLC

force which is proportional to the product of the amplitude E_0 of the driver or pump, and the amplitude x_2 of the second oscillator, the equation of motion becomes

$$\frac{d^2x_1}{dt^2} + \omega_1^2 x_1 = c_1 x_2 E_0 \tag{10.129}$$

where c_1 is a constant which indicates the strength of the mode coupling. We can write a similar equation for x_2 as follows:

$$\frac{d^2x_2}{dt^2} + \omega_2^2 x_2 = c_2 x_1 E_0 \tag{10.130}$$

Assume that $x_1 = \bar{x_1}\cos(\omega t)$, $x_2 = \bar{x_2}\cos(\omega' t)$, and $E_0 = \bar{E_0}\cos(\omega_0 t)$. Equation (10.129) reduces to

$$\begin{aligned}(\omega_2^2 - \omega'^2)\bar{x}_2\cos(\omega' t) &= c_2\bar{E}_0\bar{x}_1\cos(\omega_0 t)\cos(\omega t) \\ &= c_2\bar{E}_0\bar{x}_1(1/2)\Big[[\cos(\omega_0+\omega)t] + \cos[(\omega_0-\omega)t]\Big]\end{aligned} \tag{10.131}$$

The driving terms on the right-hand side can excite oscillators x_2 with frequencies

$$\omega' = \omega_0 \pm \omega \tag{10.132}$$

If there are no nonlinear interactions, then x_2 can have only the frequency ω_2, Thus, we have $\omega' = \omega_2$. However, the driving terms can cause a shift in the frequency, so that $\omega' \approx \omega_2$. Also, ω' can be complex because of the damping or it can grow leading to instability. In both cases, x_2 is an oscillator with finite amplitude and can respond to a range of frequencies about ω_2. If ω is small, then it is evident from equation (10.132) that both choices for ω' may lie within the bandwidth of x_2, and we must in principle make allowance for two oscillators, $x_2(\omega_0+\omega)$ and $x_2(\omega_0-\omega)$. Inserting the new variation for x_1 and x_2 as, $x_1 = \bar{x}_1\cos(\omega^\star t)$ and $x_2 = \bar{x}_2\cos[(\omega_0 \pm \omega)t]$ into equation (10.135), we have

$$\begin{aligned}(\omega_1^2-\omega^{\star 2})\bar{x}_1\cos(\omega^\star t) &= c_1\bar{E}_0\bar{x}_2(1/2)(\cos[(\omega_0+(\omega_0\pm\omega)t]+\cos[(\omega_0-(\omega_0\pm\omega)]t) \\ &= c_1\bar{E}_0\bar{x}_2(1/2)[\cos(2\omega_0 \pm \omega)t + \cos\omega t]\end{aligned} \tag{10.133}$$

In addition to the original oscillation $x_1(\omega)$, the driving terms can also excite the frequencies $\omega^\star = 2\omega_0 \pm \omega$. Consider the case when $|\omega_0| \gg |\omega_1|$, so that $2\omega_0 \pm \omega$ lies outside the range of frequencies to which x_1 would respond and neglect $x_1(2\omega_0 - \omega)$. Thus, we have coupling of the three oscillators $x_1(\omega)$, $x_2(\omega_0-\omega)$ and $x_2(\omega_0+\omega)$.

The dispersion relation is derived by setting the determinant of the coefficients equal to zero as follows:

$$\begin{bmatrix} \omega^2-\omega_1^2 & c_1E_0 & c_1E_0 \\ c_2E_0 & (\omega_0-\omega)^2-\omega_2^2 & 0 \\ c_2E_0 & 0 & (\omega_0+\omega)^2-\omega_2^2 \end{bmatrix} = 0 \tag{10.134}$$

© 2015 by Taylor & Francis Group, LLC

The solution of the above dispersion relation with $Im(\omega) > 0$ will lead to instability.

In the case of small frequency shifts and small damping or growth rates, we can set ω and ω' approximately equal to the undisturbed natural frequencies ω_1 and ω_2. Equation (10.132) gives a frequency matching condition as

$$\omega_0 \approx \omega_2 \pm \omega_1 \tag{10.135}$$

If we consider the oscillators as waves in a plasma, then we must replace ωt by $\omega t - \mathbf{k} \cdot \mathbf{r}$. The corresponding wavelength matching condition would become

$$\mathbf{k_0} = \mathbf{k_2} \pm \mathbf{k_1} \tag{10.136}$$

which will describe spatial beats. This will imply the periodicity of points of constructive and destructive interference in space. Parametric instabilities will occur at any amplitude if damping is not present. However, in practice, a small amount of either the collisional or Landau damping will prevent instability, unless the pump of the wave is very strong. In such a situation, we can introduce damping Γ_1 and Γ_2 for the oscillators x_1 and x_2 with a change in the equation as

$$\frac{d^2x_1}{dt^2} + \omega_1^2 x_1 + 2\Gamma_1 \frac{dx_1}{dt} = 0 \tag{10.137}$$

For the above case, the threshold for stability will be

$$c_1 c_2 (E_0^2)_{thresh} = 4\omega_1 \omega_2 \Gamma_1 \Gamma_2 \tag{10.138}$$

with the fact that the threshold goes to zero with the damping of either of the waves.

10.7.1 Application

Low-frequency turbulence in the solar wind is characterized by a high degree of Alfvenicity close to the Sun. Cross helicity, which is a measure of Alfvenic correlation, tends to decrease with increasing distance from the Sun at high latitudes as well as in slow-speed streams at low latitudes. However, large-scale inhomogeneities (velocity shears, the heliospheric current sheet) are present, which are sources of de correlation; moreover, at high latitudes, the wind is much more homogeneous, and a possible evolution mechanism may be represented by the parametric instability. The parametric decay of a circularly polarized broadband Alfven wave has been investigated by Malara et al.[95]. The time evolution was obtained by numerically integrating the full set of nonlinear MHD equations, up to instability saturation. They found that, for $\beta \approx 1$, the final cross-helicity is ≈ 0.5, corresponding to a partial depletion of the initial correlation. Compressive fluctuations at a moderate level are also present. Most of the spectrum is dominated by forward propagating Alfvenic fluctuations, while back scattered fluctuations dominate large scales.

© 2015 by Taylor & Francis Group, LLC

Parametric decay of circularly polarized Alfven waves with multidimensional simulations in periodic and open domains has been studied by Del Zanna et al.[96]. For higher values of beta, they found that the cross helicity decreases monotonically with time towards zero, implying an asymptotic balance between inward and outward Alfvenic modes, a feature similar to the observed decrease with distance in the solar wind. Although the instability mainly takes place along the propagation direction, in the two- and three-dimensional cases, a turbulent cascade occurs also transverse to the field. The asymptotic state of density fluctuations appears to be rather isotropic, whereas a slight preferential cascade in the transverse direction is seen in magnetic field spectra. Finally, parametric decay is shown to occur also in a nonperiodic domain with open boundaries, when the mother wave is continuously injected from one side. In two and three dimensions a strong transverse filamentation is found at long times, reminiscent of density ray-like features observed in the extended solar corona and pressure-balanced structures found in solar wind data.

The parametric interaction of beam-driven Langmuir waves in the solar wind was studied by Gurnett et al.[97], while Weatherall et al.[98] studied parametric instabilities in weakly magnetized plasmas.

10.8 Two-Stream Instability

In the following sections, we shall deal with stability properties in plasmas, such as the two stream instability, flute instability, sausage and kink instability. Finally, we shall also briefly discuss on ballooning instability.

Two-stream instability is found to occur when there are counter-streaming plasma flow in the velocity space. For our discussion, let us consider a field-free two-fluid plasma system, which consists of a cold ion fluid and a cold electron fluid. The cold ion fluid is assumed to be at rest ($V_{i0} = 0$) with uniform number density n_0. Let us assume that the electron fluid moves with a velocity $V_{e0} = V_0\hat{x}$, with number density n_0. We shall discuss the stability of this system for some electrostatic waves that propagates in the x direction. An important observation is that a two-stream plasma can lead to strong electric current in the x direction. This means that the background field should not be field free. In order to overcome this difficulty, we should consider a system with two counter-streaming electrons and one ion fluid at rest, or a system with two counter-streaming ions and one electron fluid at rest. The other option is that the system should have two counter-streaming electrons and two counter-streaming ions.

For simplicity we shall assume one ion fluid at rest and one electron fluid with velocity $V_{e0} = V_0\hat{x}$, the field structure to be two dimensional. Let $\nabla = (\partial/\partial x)\hat{x}$ in a finite extended column along the x axis. For electrostatic waves,

© 2015 by Taylor & Francis Group, LLC

we have $E_1 = -\nabla\Phi_1 = E_{x1}\hat{x}$. The linearized electrostatic two-fluid equations can be written as

$$\frac{\partial n_{i1}}{\partial t} + n_0 \frac{\partial V_{i1x}}{\partial x} = 0 \tag{10.139}$$

$$\frac{\partial n_{e1}}{\partial t} + V_0 \frac{\partial n_{e1}}{\partial x} + n_0 \frac{\partial V_{e1x}}{\partial x} = 0 \tag{10.140}$$

The above two equations are the linearized continuity equations. The linearized momentum equations can be written as

$$n_0 m_i \frac{\partial V_{i1x}}{\partial t} = e n_0 E_{1x} \tag{10.141}$$

$$n_0 m_e \left(\frac{\partial}{\partial t} + V_0 \frac{\partial}{\partial x}\right) V_{e1x} = -e n_0 E_1 \tag{10.142}$$

The Poisson equation is written as

$$\frac{\partial}{\partial x} E_1 = \frac{e}{\epsilon_0}(n_{i1} - n_{e1}) \tag{10.143}$$

Assuming a plane wave linear perturbation as $A_1(x,t) = Re\hat{A}_1(k,\omega)\exp[i(kx - \omega t)]$, the Fourier and Laplace transforms of the above equations lead to

$$-i\omega \hat{n_{i1}} + n_0 ik \hat{V_{i1x}} = 0 \tag{10.144}$$

$$-i(\omega - V_0 k)\hat{n_{e1}} + n_0 ik \hat{V_{e1x}} = 0 \tag{10.145}$$

$$n_0 m_i(-i\omega)\hat{V_{i1x}} = e n_0 \hat{E_{1x}} \tag{10.146}$$

$$n_0 m_e(-i)(\omega - V_0 k)\hat{V_{e1x}} = -e n_0 \hat{E_1} \tag{10.147}$$

$$ik\hat{E_1} = \frac{e}{\epsilon_0}(\hat{n_{i1}} - \hat{n_{e1}}) \tag{10.148}$$

Employing the usual method of determining the dispersion relation yields

$$\epsilon(k,\omega) ik \hat{E_{1x}} = 0 \tag{10.149}$$

where

$$\epsilon(k,\omega) = 1 - \left(\frac{\omega_{pi}}{\omega}\right)^2 - \left(\frac{\omega_{pe}}{\omega - kV_0}\right)^2 = 0 \tag{10.150}$$

The above expression is useful for easily determining the solution space of unstable modes. We introduce the change in variables as $x = \omega/\omega_{pe}$, $\alpha = kV_0/\omega_{pe}$. The above equation reduces to

$$1 - \frac{m_e}{m_i}\frac{1}{x^2} - \frac{1}{(x-\alpha)^2} = 0 \tag{10.151}$$

In order to determine the solution of the above equation, we shall consider the function $f(x)$ given by

$$f(x) = \frac{m_e}{m_i}\frac{1}{x^2} + \frac{1}{(x-\alpha)^2} \tag{10.152}$$

© 2015 by Taylor & Francis Group, LLC

Instability occurs when equation (10.151) has complex roots. This happens when the local minimum of $y = f(x)$ for $0 < x < \alpha$ is greater than 1. Assume that the local minimum is located at a point $x = x_A$. Then it follows that

$$f'(x_A) = -2\frac{m_e}{m_i}\frac{1}{x_A^3} - 2\frac{1}{(x_A - \alpha)^3} = 0 \tag{10.153}$$

or

$$x_A = \frac{\alpha}{1 + (m_i/m_e)^{1/3}} = \frac{\alpha}{1 + A} \approx 0.075\alpha \tag{10.154}$$

$$x^4 - 2\alpha x^3 + x^2\Big(\alpha^2 - 1 - \frac{m_e}{m_i}\Big) + \Big(2\frac{m_e}{m_i}\alpha\Big)x - \alpha^2\frac{m_e}{m_i} = 0 \tag{10.155}$$

where $A = (m_i/m_e)^{1/3}$.

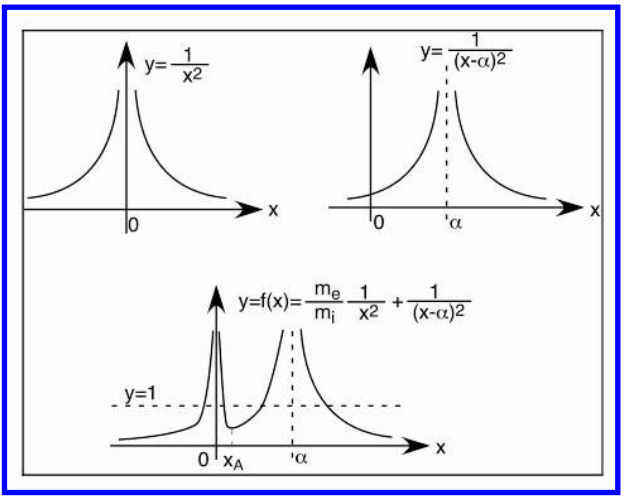


FIGURE 10.9: Plots of the three functions for determining the stability criterion, from Lyu [24].

Figure 10.9 presents the graph of the following functions: (a) $y = 1/x^2$, (b) $y = 1/(x-\alpha)^2$, and (c) $y = f(x) = \frac{m_e}{m_i}\frac{1}{x^2} + \frac{1}{(x-\alpha)^2}$, respectively. It is very easy to realize that the two-stream instability occurs when $f(x = x_A) > 1$. The analytical expression for the instability condition becomes

$$f(x_A) = \frac{m_e}{m_i}\frac{1}{x_A^2} + \frac{1}{(x_A - \alpha)^2} = \frac{(1+A)^2}{A^3\alpha^2} + \frac{(1+A)^2}{\alpha^2 A^2} = \frac{(1+A)^3}{\alpha^2 A^3} > 1 \tag{10.156}$$

Simple algebra yields the value for alpha to be $\alpha < 1.12486$ or equivalently $kV_0 < 1.12486\ \omega_{pe}$. We can rewrite the dispersion relation as

$$x^2(x-\alpha)^2 - x^2 - \frac{m_e}{m_i}(x-\alpha)^2 = 0 \tag{10.157}$$

© 2015 by Taylor & Francis Group, LLC

which can be simplified as

$$x^4 - 2\alpha x^3 + x^2\Big(\alpha^2 - 1 - \frac{m_e}{m_i}\Big) + \Big(2\frac{m_e}{m_i}\alpha\Big)x - \alpha^2\frac{m_e}{m_i} = 0 \qquad (10.158)$$

Figure 10.10 presents all solutions of ω as a function of the wavenumber k, which include one real root for $\omega > \alpha$ and one real root when $\omega < 0$. The other two roots are real, given by $\omega = \omega_{r1}$ and ω_{r2}. The two complex roots are conjugate in nature given by $\omega = \omega_r \pm i\omega_i$. The most unstable mode occurs near $\alpha \approx 1$ or $kV_0 \approx \omega_{pe}$, as is evident from the lower panel of Figure 10.10.

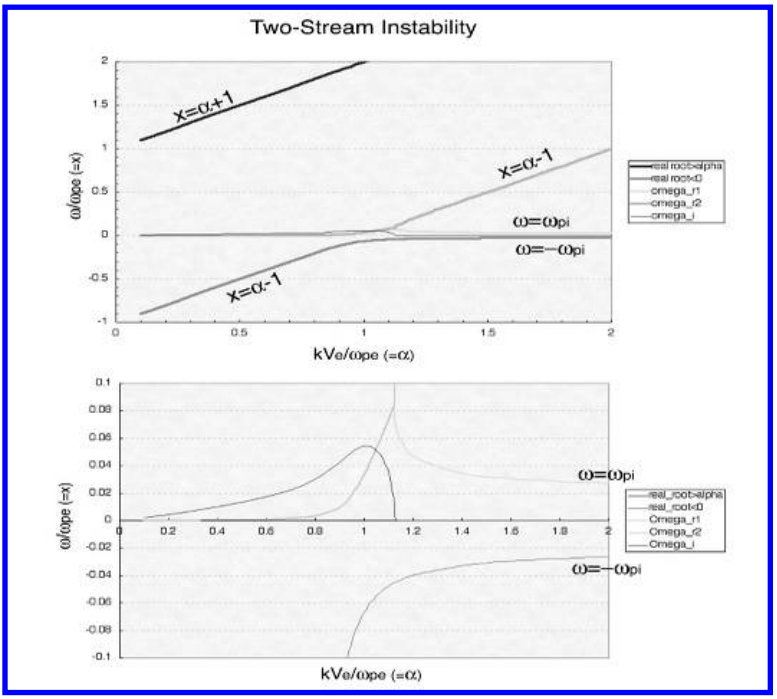


FIGURE 10.10: Solutions of Equation (10.157) plotted with x as a function of α, from Lyu [24].

By simple algebra, we can show that for $\omega_{pi} \to 0$, the four roots are given by $x = 0, 0, \alpha + 1$, and $\alpha - 1$, or $\omega = 0, 0, kV_0 + \omega_{pe}$, and $kV_0 - \omega_{pe}$. Similarly for the finite ion mass, the four wave modes are

$$x = \omega_{pi}/\omega_{pe}, -\omega_{pi}/\omega_{pe}, \alpha + 1, \alpha - 1 \qquad (10.159)$$

It can be seen from the top panel of Figure 10.10 that the four real roots at short wavelength limit ($kV_0/\omega_{pe} \gg 1$) approach the solutions as mentioned above. However, in the long wavelength limit where $kV_0/\omega_{pe} < 1$, two real roots approach $x = \alpha + 1$ and $\alpha - 1$ or $\omega = kV_0 + \omega_{pe}$ and $kV_0 - \omega_{pe}$.

© 2015 by Taylor & Francis Group, LLC

The maximum growth rate occurs near the intersection of $x = \omega_{pi}/\omega_{pe}$ and $x = \alpha - 1$. After the Doppler shift, the wave mode which is close to the ion's plasma frequency becomes electrons' plasma frequency in the electrons moving frame.

10.9 Interchange (Flute) Instability

Let $x = 0$ be the boundary separating a plasma and vacuum and assume that the z axis is in the direction of the magnetic field $\mathbf{B}$. The plasma region is $x < 0$ and the vacuum region is $x > 0$. The acceleration due to gravity g is in the x direction. Both ions and electrons drift in opposite directions to each other, due to the acceleration, with drift velocities given by

$$v_{G,i} = \frac{M}{e}\frac{g \times \mathbf{B}}{B^2} \tag{10.160}$$

$$v_{G,e} = -\frac{m}{e}\frac{g \times \mathbf{B}}{B^2} \tag{10.161}$$

Let us also assume that, due to a perturbation, the boundary of the plasma is displaced from the surface $x = 0$, by an amount given by

$$\delta x = a(t)\sin(k_y y) \tag{10.162}$$

The charge separation due to the opposite ion and electron drifts yields an electric field. The resultant $\mathbf{E} \times \mathbf{B}$ drift enhances the original perturbation if the direction of the acceleration g is outward from the plasma. It is easy to see that this is the same as saying that the magnetic flux originally inside but near the plasma boundary is displaced so that it is outside the boundary, while the flux outside moves in to fill the depression thus left in the boundary. Because of this geometrical picture of the process, this type of instability has come to be called interchange instability. As the perturbed plasma boundary is in the form of flutes along the magnetic lines of force, this instability is also called flute instability. The drift in the ions and electrons, resulting in the electric field for interchange instability is shown in Figure 10.11.

The drift due to the acceleration yields a surface charge on the plasma, of charge density

$$\sigma_s = \sigma(t)\cos(k_y y)\delta(x) \tag{10.163}$$

The electrostatic potential ϕ of the induced electric field $\mathbf{E} = -\nabla\phi$ is given by

$$\epsilon_\perp \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial}{\partial x}\left(\epsilon_\perp \frac{\partial \phi}{\partial x}\right) = -\sigma_s \tag{10.164}$$

© 2015 by Taylor & Francis Group, LLC

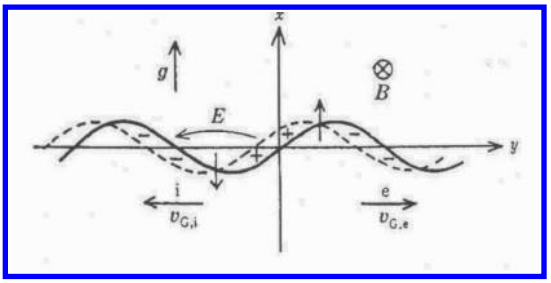


FIGURE 10.11: Ion and electron drifts and the resultant electric field for interchange instability, from Miyomoto [20].

The boundary condition is given by

$$\epsilon_0\Big(\frac{\partial\phi}{\partial x}\Big)_{+0} - \Big(\epsilon_\perp\frac{\partial\phi}{\partial x}\Big)_{-0} = -\sigma_s \tag{10.165}$$

$$\phi_{+0} = \phi_{-0} \tag{10.166}$$

If we assume that $k_y > 0$, then the solution for ϕ can be immediately written as

$$\phi = \frac{\sigma(t)}{k_y(\epsilon_0+\epsilon_\perp)}\cos(k_y y)\exp(-k_y|x|) \tag{10.167}$$

The velocity of the boundary $d(\delta x)/dt$ is equal to $\mathbf{E}\times\mathbf{B}/B^2$ at $x = 0$, with $\mathbf{E}$ being determined from the potential. The velocity can be calculated to yield

$$\frac{da(t)}{dt}\sin(k_y y) = \frac{\sigma(t)}{(\epsilon_0+\epsilon_\perp)B}\sin(k_y y) \tag{10.168}$$

The charge flux in the y direction is given by

$$ne|v_{G,i}| = \frac{\rho_m g}{B} \tag{10.169}$$

where $\rho_m = nM$. The rate of change of charge density is given by

$$\frac{d\sigma(t)}{dt}\cos(k_y y) = \frac{\rho_m g}{B}a(t)\frac{d}{dt}\sin(k_y y) \tag{10.170}$$

and

$$\frac{d^2a}{dt^2} = \frac{\rho_m g k_y}{(\epsilon_0+\epsilon_\perp)B^2}a \tag{10.171}$$

The solution is in the form $a \propto \exp(\gamma t)$, where γ is the growth rate, which is given by

$$\gamma = \Big(\frac{\rho_m}{(\epsilon_0+\epsilon_\perp)B^2}\Big)^{1/2}(gk_y)^{1/2} \tag{10.172}$$

© 2015 by Taylor & Francis Group, LLC

For the low frequency case (as compared to the ion cyclotron frequency), the dielectric constant is given by

$$\epsilon_\perp = \epsilon_0\left(1 + \frac{\rho_m}{B^2\epsilon_0}\right)^{1/2} \gg \epsilon_0 \tag{10.173}$$

The growth rate in this case is given by

$$\gamma = (gk_y)^{1/2} \tag{10.174}$$

For the case when the acceleration is outward, i.e., a perturbation with the propagation vector $\mathbf{k}$ normal to the magnetic field $\mathbf{B}$, is unstable:

$$(\mathbf{k}\cdot\mathbf{B}) = 0 \tag{10.175}$$

In contrast, if the acceleration is inward ($g < 0$), then the growth rate γ is imaginary and so the perturbation is oscillatory and stable. The origin of interchange instability is charge separation due to the acceleration. When the magnetic lines of force are curved, the charged particles are subjected to a centrifugal force. If the magnetic lines of force are convex outward, this centrifugal acceleration induces interchange instability. If the lines are concave outward, the plasma is stable. Accordingly, the plasma is stable when the magnitude B of the magnetic field increases outward. In other words, if B is a minimum at the plasma region, the plasma is stable. This is the minimum-B condition for stability.

The drift motion of the charged particles is expressed by

$$v_g = \frac{\mathbf{E}\times\mathbf{b}}{B} + \frac{\mathbf{b}}{\Omega}\times\left(g + \frac{(v_\perp^2/2) + v_\parallel^2}{R}\mathbf{n}\right) + v_\parallel\mathbf{b} \tag{10.176}$$

where $\mathbf{n}$ is the normal unit vector from the center of curvature to a point on a line of magnetic force. R is the radius of curvature of the line of magnetic force.

The equivalent acceleration is given by

$$g = \frac{(v_\perp^2/2) + v_\parallel^2}{R}\mathbf{n} \tag{10.177}$$

The growth rate in this case becomes $\gamma \approx (a/R)^{1/2}(vT/a)$. For a perturbation with propagation vector $\mathbf{k}$ normal to the magnetic field $\mathbf{B}$, i.e., $(\mathbf{k}\cdot\mathbf{B}) = 0$, another mechanism of charge separation may cause the same type of instability. When a plasma rotates with the velocity $v_\theta = E_r/B$ due to an inward radial electric field, and if the rotation velocity of ions falls below that of electrons, the perturbation is unstable. Several possible mechanisms can retard ion rotation. The collision of ions and neutral particles delays the ion velocity and causes neutral drag instability.

© 2015 by Taylor & Francis Group, LLC

When the growth rate $\gamma \approx (gk_y)^{1/2}$ is not very large and the ion Larmor radius ρ_Ω^i is large enough to satisfy

$$(k_y \rho_\Omega^i)^2 > \frac{\gamma}{|\Omega_i|} \tag{10.178}$$

the perturbation is stabilized. When the ion Larmor radius becomes large, the average perturbation electric field felt by the ions is different from that felt by the electrons, and the $\mathbf{E} \times \mathbf{B}/B^2$ drift velocities of the ion and the electrons are different. The charge separation thus induced has opposite phase from the charge separation due to acceleration and stabilizes the instability.

10.10 Sausage Instability

Consider a cylindrical plasma with a sharp boundary. Assume that a longitudinal magnetic field B_z exists inside the plasma region and an azimuthal field $H_\theta = I_z/2\pi r$, due to the plasma current I_z, exists outside the plasma region. In what follows, we shall examine the azimuth-ally symmetric perturbation which constricts the plasma like a sausage (see Figure 10.12).

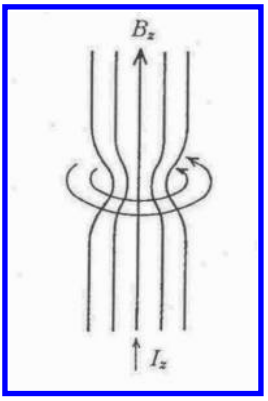


FIGURE 10.12: A diagram for sausage instability, from Miyomoto[20].

When the plasma radius a is changed to δa, the conservation of magnetic flux and the current in the plasma yields the following relations

$$\delta B_z = -B_z \frac{2\delta a}{a} \tag{10.179}$$

$$\delta B_\theta = -B_\theta \frac{\delta a}{a} \tag{10.180}$$

The longitudinal magnetic field inside the plasma acts against the perturbation, while the external azimuthal field destabilizes the perturbation. The

© 2015 by Taylor & Francis Group, LLC

difference δp_m in the magnetic pressures is

$$\delta p_m = -\frac{B_z^2}{\mu_0}\frac{2\delta a}{a} + \frac{B_\theta^2}{\mu_0}\frac{\delta}{a} \tag{10.181}$$

The plasma is stable if $\delta p_m > 0$ for $\delta < 0$, so that the stability condition can be written as

$$B_z^2 > \frac{B_\theta^2}{2} \tag{10.182}$$

The above expression represents an instability called the sausage instability. In the next section, we shall briefly talk about kink instability.

10.11 Kink Instability

Let us consider a perturbation that kinks the plasma column as shown in Figure 10.13. The configuration of the plasma is the same as that in the previous subsection (sharp boundary, an internal longitudinal field, an external azimuthal field). Let us denote the characteristic length of the kink by λ and its radius of curvature by R. The longitudinal magnetic field acts as a restoring force on the plasma due to the longitudinal tension; the restoring force on the plasma region of length λ is

$$\frac{B_z^2}{2\mu_0}\pi a^2\frac{\lambda}{R} \tag{10.183}$$

The azimuthal magnetic field becomes strong at the inner (concave) side of the kink and destabilizes the plasma column. In order to estimate the destabilizing force, we consider a cylindrical lateral surface of radius λ around the plasma and two planes A and B which pass through the center of curvature (see Figure 10.13).

Let us compare the contributions of the magnetic pressure on the surfaces enclosing the kink. The contribution of the magnetic pressure on the cylindrical surface is negligible compared with those on the planes A and B. The contribution of the magnetic pressure on the planes A and B is

$$\int_a^\lambda \frac{B_\theta^2}{2\mu_0}2\pi r dr \times \frac{\lambda}{2R} = \frac{B_\theta^2(a)}{2\mu_0}\pi a^2 \ln\frac{\lambda}{a} \times \frac{\lambda}{R} \tag{10.184}$$

The above expression yields the simple stability condition as

$$\frac{B_z^2}{B_\theta^2(a)} > \ln\frac{\lambda}{a} \tag{10.185}$$

It is important to realize that the pressure balance relation which is

$$p + \frac{B_z^2}{2\mu_0} = \frac{B_\theta^2}{2\mu_0} \tag{10.186}$$

© 2015 by Taylor & Francis Group, LLC

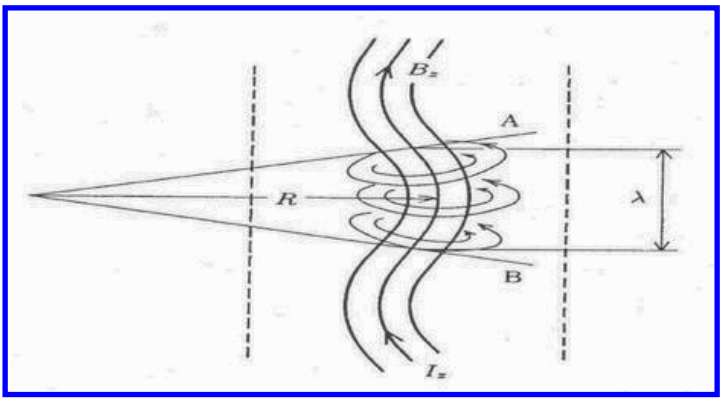


FIGURE 10.13: A simple sketch for kink instability, from Miyomoto [20].

still holds, so that perturbations with large value of λ are unstable. This type of instability is termed the kink instability.

10.12 Ballooning Instability

In the case of interchange instability, the parallel component $k_\parallel = (\mathbf{k}\cdot\mathbf{B}/B)$ of the propagation vector is zero and the minimum-B condition may stabilize such an instability. In this section, we shall study perturbations wherein $k_\parallel \neq 0$, but $|k_\parallel/k_\perp| \ll 1$. Although the interchange instability may be stabilized by an average minimum-B configuration, it is possible that the perturbation with $k_\parallel \neq 0$ can grow locally in a region of the average minimum-B field. This type of instability is called the ballooning instability.

The energy integral δW is given by

$$\begin{aligned}\delta W &= \frac{1}{2\mu_0}\int((\nabla\times(\xi\times\mathbf{B_0}))^2 - (\xi\times(\nabla\times\mathbf{B_0}))\cdot\nabla\times(\xi\times\mathbf{B_0}) \\ &\quad +\gamma\mu_0 p_0(\nabla\cdot\xi)^2 + \mu_0(\nabla\cdot\xi)(\xi\cdot\nabla p_0))d\mathbf{r} \qquad (10.187)\end{aligned}$$

Let us consider the case when ξ can be expressed as

$$\xi = \frac{\mathbf{B_0}\times\nabla\phi}{B_0^2} \qquad (10.188)$$

where ϕ may be considered to be the time integral of the scalar electrostatic potential of the perturbed electric field. One can easily check that

$$\xi\times\mathbf{B_0} = \nabla_{\perp\phi} \qquad (10.189)$$

© 2015 by Taylor & Francis Group, LLC

With this, the energy integral can be simplified to yield

$$\begin{aligned}\delta W &= \frac{1}{2\mu_0}\int\Big((\nabla\times\nabla_{\perp\phi})^2-\Big(\frac{\mathbf{B_0}\times\nabla_{\perp\phi}\times\mu_0 j_0}{B_0^2}\Big)\nabla\times\nabla_{\perp\phi}\\ &\quad+\gamma\mu_0 p_0(\nabla\cdot\xi)^2+\mu_0(\nabla\cdot\xi)(\xi\cdot\nabla p_0)\Big)d\mathbf{r}\end{aligned}\tag{10.190}$$

Here $\nabla\cdot\xi$ is given by

$$\begin{aligned}\nabla\cdot\xi &= \nabla\cdot\Big(\frac{\mathbf{B_0}\times\nabla\phi}{B_0^2}\Big)=\nabla\phi\cdot\nabla\times\Big(\frac{\mathbf{B_0}}{B_0^2}\Big)\\ &= \nabla\phi\cdot\Big[\Big(\nabla\frac{1}{B^2}\Big)\times\mathbf{B}+\frac{1}{B^2}\nabla\times\mathbf{B}\Big]\end{aligned}\tag{10.191}$$

Defining $\nabla p_0=\mathbf{j_0}\times\mathbf{B_0}$, δW can be written as

$$\begin{aligned}\delta W &= \frac{1}{2\mu_0}\int(\nabla\times\nabla_{\perp\phi})^2+\frac{\mu_0\nabla p_0\cdot(\nabla_{\perp\phi}\times\mathbf{B_0})}{B_0^2}\Big(\frac{\mathbf{B_0}\cdot\nabla\times\nabla_{\perp\phi}}{B_0^2}\Big)\\ &\quad-\frac{\mu_0(\mathbf{j_0}\cdot\mathbf{B_0})}{B_0^2}\nabla_{\perp\phi}\cdot\nabla\times\nabla_{\perp\phi}+\gamma\mu_0 p_0\Big[\nabla\Big(\frac{1}{B_0^2}\Big)\cdot(\mathbf{B_0}\times\nabla_{\perp\phi})\Big]^2\\ &\quad+\frac{\mu_0\nabla p_0\cdot(\mathbf{B_0}\times\nabla_{\perp\phi})}{B_0^2}\Big[\nabla\Big(\frac{1}{B_0^2}\Big)\cdot(\mathbf{B_0}\times\nabla_{\perp\phi})\Big]d\mathbf{r}\end{aligned}\tag{10.192}$$

Consider the z coordinate as a length along a field line r as the radial coordinate of the magnetic surfaces and θ as the poloidal angle in the perpendicular direction to the field lines. The (r,θ,z) components of ∇p_0, $\mathbf{B}$, and $\nabla\phi$ are approximately given as

$$\nabla p_0=(p_0',0,0),\qquad \mathbf{B}=(0,B_\theta(r),B_0(1-rR_c^{-1}(z)))$$

$$\nabla\phi=(\partial\phi/\partial r,\partial\phi/r\partial\theta,\partial\phi/\partial z),\qquad \phi(r,\theta,z)=\phi(r,z)Re[e^{im\theta}]$$

$R_c(z)$ is the radius of curvature of the line of magnetic force given by

$$\frac{1}{R_c(z)}=\frac{1}{R_0}\Big(-w+\cos 2\pi\frac{z}{L}\Big)$$

For the case when the configuration is average minimum-B, w and R_0 must be $1>w>0$ and $R_0>1$. Assuming that B_θ/B_0, and r/R_0, r/L are all small quantities, we can approximate them as

$$\nabla_{\perp\phi}=\nabla\phi-\nabla_{\|\phi}\approx Re\Big(\frac{\partial\phi}{\partial r},\frac{im}{r}\phi,0\Big)$$

$$\nabla\times(\nabla_{\perp\phi})\approx Re\Big(\frac{-im}{r}\frac{\partial\phi}{\partial z},\frac{\partial^2\phi}{\partial z\partial r},0\Big)$$

$$\mathbf{B_0}\times\nabla_{\perp\phi}\approx Re\Big(\frac{-im}{r}B_0\phi,B_0\frac{\partial\phi}{\partial r},0\Big)$$

© 2015 by Taylor & Francis Group, LLC

so that δW can be written as

$$\delta W = \frac{1}{2\mu_0}\int \frac{m^2}{r^2}\Big[\Big(\frac{\partial\phi(r,z)}{\partial z}\Big)^2 - \frac{\beta}{r_p R_c(z)}(\phi(r,z))^2\Big]2\pi r dr dz$$

where $-p_0/p_0' = r_p$ and $\beta = p_0/(B_0^2/2\mu_0)$. The second term contributes to the stability in the region $R_c(z) < 0$ and leads to instability in the region $R_c(z) > 0$.

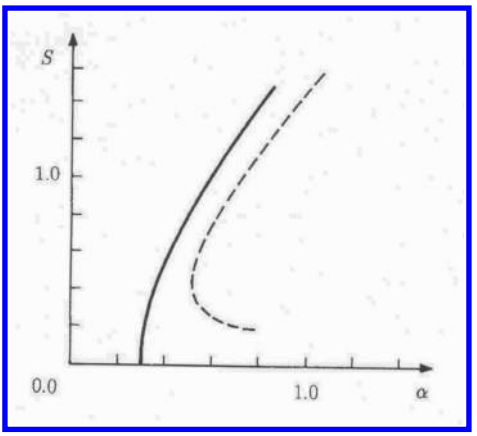


FIGURE 10.14: The maximum stable pressure gradient α as a function of the shear parameter S of the ballooning mode, from Miyomoto [20].

The stable region in the shear parameter S and the maximum pressure gradient α of the ballooning mode is shown in Figure 10.14. The shear parameter and the pressure gradient are defined as follows:

$$S = \frac{r}{q}\frac{dq}{dr}$$

and

$$\alpha = -\frac{q^2 R}{B^2/2\mu_0}\frac{dp}{dr}$$

where q is a measure of the rotational transform angle. It must be noted that the ballooning mode is stable in the negative shear region of S, as is shown in Figure 10.14. When the shear parameter S is negative ($q(r)$ decreases outwardly), the outer lines of magnetic force rotate around the magnetic axis more quickly than the inner ones.

In this book, we have tried to discuss waves and oscillations in different types of systems, with some applications to geophysics and astrophysics. We have concentrated mostly on the linear aspects of the theory. However, wherever possible, we have discussed some nonlinear aspects also. The book is far from complete. Any interested reader who wishes to know more details on the

© 2015 by Taylor & Francis Group, LLC

topics covered should take time to look into the bibliography. We hope that this book will serve as a starting point for researchers in the area of theoretical physics.

10.13 Exercises

1. The Rayleigh stability equation is

$$(U - c)(\phi'' - k^2\phi) - U''\phi = 0$$

For a velocity distribution $U = \pm U_0\hat{z}$, where U_0 is a constant, for $z >$ *and* < 0, respectively, solve for ϕ with the boundary conditions $k\phi = 0$ at $z = 0$ and $z = z_1$.

2. Solve the Taylor−Goldstein equation for a tangential velocity discontinuity $U(z) = \pm U_0\hat{z}$ for $z >$ and < 0.

3. Assume the basic shear is such that $U'' = 0$. Derive the condition for the solution of the wave equation to have a sinusoidal solution.

4. Under what condition does the Orr−Sommerfeld equation reduce to the Taylor−Goldstein equation?

5. The dispersion relation in dealing with Rayleigh – Taylor instability for an incompressible fluid is given by $\gamma^2 = k\ g\ \tan h[k_\perp h]$. Discuss the stability of this configuration. Discuss possible stabilizing effects.

6. What should be the condition for stability when the magnetic field B_0 is introduced?

7. If α is the angle between the gradient of shear and $\hat{k}$, while θ_1 and θ_2 are the angles between the magnetic fields B_1 and B_2, show that for instability

$$\triangle U^2 \cos^2\alpha > \frac{1}{\mu_0}\Big(\frac{1}{\rho_1} + \frac{1}{\rho_2}\Big)(B_1^2\cos^2\theta_1 + B_2^2\cos^2\theta_2)$$

8. The dispersion relation for discussion on two-stream instability is written as (see section on two-stream instability)

$$1 - \frac{m_e}{m_i}\frac{1}{x^2} - \frac{1}{(x-\alpha)^2} = 0$$

Derive the expression for local minimum which leads to instability.

© 2015 by Taylor & Francis Group, LLC

9. The growth rate for the flute instability is $\gamma = (gk_y)^{1/2}$. When is γ imaginary? Discuss the consequences.

10. Discuss the important difference between the sausage instability and the kink instability.

© 2015 by Taylor & Francis Group, LLC

Appendix A

Typical Tables

TABLE A.1: Comparison between Gravitation and Electrostatics

Gravity	Electrostatics
Mass; m	Charge; q
Gravitation force: $\mathbf{F}_G = -G\dfrac{mM}{r^2}\hat{\mathbf{r}}$	Electric force: $\mathbf{F}_e = -k\dfrac{qQ}{r^2}\hat{\mathbf{r}}$
Gravitation field: $\mathbf{g} = \dfrac{\mathbf{F}_G}{m}$	Electric field: $\mathbf{E} = \dfrac{\mathbf{F}_e}{q}$
Potential energy change: $\Delta U = -\int_{P_1}^{P_2} \mathbf{F}_G \cdot d\mathbf{l}$	Potential energy change: $\Delta U = -\int_{P_1}^{P_2} \mathbf{F}_e \cdot d\mathbf{l}$
Gravitational potential: $\Delta V_G = -\int_{P_1}^{P_2} \mathbf{g} \cdot d\mathbf{l}$	Electrical potential: $\Delta V = -\int_{P_1}^{P_2} \mathbf{E} \cdot d\mathbf{l}$

TABLE A.2: Static Electric Field Relations

	Definition	Units
Coulomb's law	$\mathbf{F} = \dfrac{qQ}{4\pi\epsilon r^2}\hat{\mathbf{r}}$	newtons (N)
Electric field	$\mathbf{E} = \dfrac{\mathbf{F}}{q}$	$\mathrm{V.m^{-1}}$ or $\mathrm{N.C^{-1}}$
Flux density	$\mathbf{D} = \epsilon\mathbf{E}$	$\mathrm{C.m^{-2}}$
Gauss' law	$\oint_S \mathbf{D} \cdot d\mathbf{S} = Q_{enclosed}$	C
Electric flux	$\Phi_E = \int_S \mathbf{D} \cdot d\mathbf{S}$	C
Potential difference	$V_{21} = -\int_{P_1}^{P_2} \mathbf{E} \cdot d\mathbf{l}$	V or $\mathrm{J.C^{-1}}$

© 2015 by Taylor & Francis Group, LLC

TABLE A.3: Static Magnetic Field Relations

	Definition	Units
Biot-Savart law	$d\mathbf{H} = \dfrac{id\mathbf{l}\times\hat{\mathbf{r}}}{4\pi R^2}$	$\mathrm{A.m}^{-1}$
Flux density	$\mathbf{B} = \mu\mathbf{H}$	T
Magnetic flux	$\Phi_B = \int_{\mathsf{S}} \mathbf{B}\cdot d\mathbf{S}$	Wb
Ampère's law	$\oint \mathbf{H}\cdot d\mathbf{l} = i_{enclosed}$	A
Lorentz force	$\mathbf{F} = i\mathbf{L}\times\mathbf{B}$	N

TABLE A.4: Maxwell's Equations

Differential form	Integral form
$\boldsymbol{\nabla}\times\mathbf{E}(\mathbf{r},t) = -\dfrac{\partial\mathbf{B}(\mathbf{r},t)}{\partial t}$	$\oint_{\mathsf{C}} \mathbf{E}(\mathbf{r},t)\cdot d\mathbf{l} = -\int_{\mathsf{S}} \dfrac{\partial\mathbf{B}(\mathbf{r},t)}{\partial t}\cdot d\mathbf{S}$
$\boldsymbol{\nabla}\times\mathbf{H}(\mathbf{r},t) = \mathbf{J} + \dfrac{\partial\mathbf{D}(\mathbf{r},t)}{\partial t}$	$\oint_{\mathsf{C}} \mathbf{H}(\mathbf{r},t)\cdot d\mathbf{l} = \int_{\mathsf{S}} \left(\mathbf{J}(\mathbf{r},t) + \dfrac{\partial\mathbf{D}(\mathbf{r},t)}{\partial t}\right)\cdot d\mathbf{S}$
$\boldsymbol{\nabla}\cdot\mathbf{D}(\mathbf{r},t) = \rho(\mathbf{r},t)$	$\oint_{\mathsf{S}} \mathbf{D}(\mathbf{r},t)\cdot d\mathbf{S} = \int_{\mathsf{V}} \rho(\mathbf{r},t)d\mathsf{V}$
$\boldsymbol{\nabla}\cdot\mathbf{B}(\mathbf{r},t) = 0$	$\oint_{\mathsf{S}} \mathbf{B}(\mathbf{r},t)\cdot d\mathbf{S} = 0$

© 2015 by Taylor & Francis Group, LLC

Appendix B

Vector Operators

B.1 Vector Formulae 495
B.1.1 Scalar Product 496
B.1.2 Vector Product 497
B.1.3 Del Operator 497
B.1.4 Laplacian Operator 498
B.2 Vector Derivatives 498
B.2.1 Gradient 499
B.2.2 Divergence 499
B.2.3 Laplacian 500
B.2.4 Curl 500
B.3 Stokes Theorem 501
B.4 Green's Theorem 501

B.1 Vector Formulae

The physical quantities are divided into scalars and vectors according to the range of information denoted by the quantity. The scalar quantity, for instance, speed, time, distance, pressure, and temperature, has magnitude only. On the other hand the vector quantity, for example, acceleration, velocity, force, and displacement, has both magnitude and direction. A displacement vector from a point A to another point B is denoted by $\mathbf{AB}$. In what follows, we elucidate some important definitions briefly:

1. Unit vector: A unit vector is a vector having a magnitude of one unit and points in a particular direction. The unit vector along any vector, for example, $\mathbf{A}(= A_x, A_y, A_z)$, in which A_x, A_y, and A_z are the magnitudes of x, y and z components of $\mathbf{A}$ respectively, is calculated as

$$\hat{\mathbf{A}} = \frac{\mathbf{A}}{|\mathbf{A}|}$$

Three unit vectors in the direction of x, y, and z axes are, in general, used, which are respectively given by $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$. A vector is denoted by using the concept of unit vector and components of vector in the unit vector form as

$$\mathbf{A} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}$$

2. Magnitude (or length) of a vector: The magnitude of a vector can be

© 2015 by Taylor & Francis Group, LLC

found using the Pythagorean theorem. The magnitude of a vector $\mathbf{A}$ is denoted by

$$|\mathbf{A}| = \sqrt{A_x^2 + A_y^2 + A_z^2}$$

3. Position vectors: The position vectors provide the position of a point, (x, y, z), relative to a fixed point (the origin):

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

4. Vector components: The components of a vector are two or more vectors along the x, y, and z axes whose vector sum is equal to the given vector. $\mathbf{AB} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$ and the distance between the two points is

$$|\mathbf{AB}| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

5. Vector addition: If two forces, $\mathbf{A}$ and $\mathbf{B}$, are acting in a same direction, the resultant force $\mathbf{R}$ becomes the sum of these vectors, that is,

$$\mathbf{R} = \mathbf{A} + \mathbf{B}$$

6. Vector subtraction: If two forces, $\mathbf{A}$ and $\mathbf{B}$ are acting in directions opposite each other, the resultant force $\mathbf{R}$ becomes the sum of these vectors, that is,

$$\mathbf{R} = \mathbf{A} - \mathbf{B}$$

There are two ways to take the product of a pair of vectors, which are enumerated below briefly.

B.1.1 Scalar Product

The scalar (or dot) product of two vectors, $\mathbf{A}$ and $\mathbf{B}$, is written as

$$\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos\theta \tag{B.1}$$

in which $|\mathbf{A}|$ and $|\mathbf{B}|$ are the magnitudes of vectors $\mathbf{A}$ and $\mathbf{B}$, respectively, and θ the angle between these vectors; two vectors are perpendicular (orthogonal) if their dot product is zero.

Expressed in terms of Cartesian components of the two vectors, the dot product is given by

$$\mathbf{A}\cdot\mathbf{B} = A_x B_x + A_y B_y + A_z B_z \tag{B.2}$$

To note, since the standard unit vectors are orthogonal, the dot product between a pair of distinct standard unit vectors is zero:

$$\mathbf{i}\cdot\mathbf{j} = \mathbf{i}\cdot\mathbf{k} = \mathbf{j}\cdot\mathbf{k} = 0$$

In the case of the angle being 0, the dot products are

$$\mathbf{i}\cdot\mathbf{i} = \mathbf{j}\cdot\mathbf{j} = \mathbf{k}\cdot\mathbf{k} = 1$$

© 2015 by Taylor & Francis Group, LLC

B.1.2 Vector Product

The vector (cross) product is defined only for three-dimensional (3-D) vectors. If $\mathbf{A}$ and $\mathbf{B}$ are 3-D vectors, the cross product is given by

$$\mathbf{A}\times\mathbf{B} = |\mathbf{A}||\mathbf{B}|\sin\theta \tag{B.3}$$

in which θ is the angle between $\mathbf{A}$ and $\mathbf{B}$.

This formula (equation B.3) shows that the magnitude of the cross product is largest when $\mathbf{A}$ and $\mathbf{B}$ are perpendicular, while if these vectors are parallel or if either vector is the zero vector, the cross product turns out to be the zero vector. We define this cross product vector by the following requirements:

1. $\mathbf{A}\times\mathbf{B}$ is perpendicular to both $\mathbf{A}$ and $\mathbf{B}$ with the orientation determined by the right-hand rule.
2. The magnitude of the vector $\mathbf{A}\times\mathbf{B}$ expressed as $|\mathbf{A}\times\mathbf{B}|$ is the area of the parallelogram spanned by $\mathbf{A}$ and $\mathbf{B}$.

The cross product can be written as

$$\mathbf{A}\times\mathbf{B} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \tag{B.4}$$

where $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ are unit vectors.

In three-dimensional Cartesian coordinates, these unit vectors are

$$\hat{\mathbf{i}}\times\hat{\mathbf{j}} = \hat{\mathbf{k}}, \qquad \hat{\mathbf{j}}\times\hat{\mathbf{k}} = \hat{\mathbf{i}}, \qquad \hat{\mathbf{k}}\times\hat{\mathbf{i}} = \hat{\mathbf{j}}$$

The cross product satisfies the general identity

$$\mathbf{A}\times\mathbf{B} = -\mathbf{B}\times\mathbf{A} \tag{B.5}$$

The identities involving cross product include:

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A}) = \mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}) \tag{B.6}$$

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = (\mathbf{A}\cdot\mathbf{C})\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\mathbf{C} \tag{B.7}$$

$$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D}) = (\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D}) - (\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}) \tag{B.8}$$

B.1.3 Del Operator

The del operator is a differential quantity, which can be defined in different coordinate systems as the following:

1. Cartesian coordinates:

$$\boldsymbol{\nabla} = \hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z} \tag{B.9}$$

© 2015 by Taylor & Francis Group, LLC

2. Cylindrical polar coordinates:

$$\boldsymbol{\nabla} = \frac{\partial}{\partial \rho}\hat{\boldsymbol{\rho}} + \frac{1}{\rho}\frac{\partial}{\partial \phi}\hat{\boldsymbol{\phi}} + \frac{\partial}{\partial z}\hat{\mathbf{k}} \tag{B.10}$$

3. Spherical polar coordinates:

$$\boldsymbol{\nabla} = \frac{\partial}{\partial r}\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial}{\partial \theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\hat{\boldsymbol{\phi}} \tag{B.11}$$

in which the spherical coordinates r, θ, and ϕ are related to the Cartesian coordinates x, y, and z by

$$x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi, \quad z = r\cos\theta \tag{B.12}$$

and

$$r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2} \tag{B.13}$$

B.1.4 Laplacian Operator

The Laplacian operator is a second order differential quantity, which can be defined as

1. Cartesian coordinates:

$$\boldsymbol{\nabla}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{B.14}$$

2. Cylindrical polar coordinates:

$$\boldsymbol{\nabla}^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2} \tag{B.15}$$

3. Spherical polar coordinates:

$$\boldsymbol{\nabla}^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial \phi^2} \tag{B.16}$$

B.2 Vector Derivatives

Vector derivatives are used in several branches of physics, for example, fluid mechanics, electricity and magnetism, elasticity, and many other areas of theoretical and applied physics.

© 2015 by Taylor & Francis Group, LLC

B.2.1 Gradient

Gradient is a vector operation operating on a scalar function to produce a vector whose magnitude is the maximum rate of change of the function at the point of the gradient and which is pointed in the direction of that maximum rate of change. In Cartesian coordinates, the gradient of function $f(x, y, z)$ is

$$\boldsymbol{\nabla} f = \left[\hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z}\right] f \tag{B.17}$$

In cylindrical polar coordinates, the gradient of a scalar is

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial \rho}\hat{\boldsymbol{\rho}} + \frac{1}{\rho}\frac{\partial f}{\partial \phi}\hat{\boldsymbol{\phi}} + \frac{\partial}{\partial z}\hat{\mathbf{k}} \tag{B.18}$$

and in spherical polar coordinates, it is

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial r}\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial f}{\partial \theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\hat{\boldsymbol{\phi}} \tag{B.19}$$

B.2.2 Divergence

The divergence in Cartesian coordinates is defined as

$$\boldsymbol{\nabla}\cdot\mathbf{A} = \left(\hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial k} + \hat{\mathbf{z}}\frac{\partial}{\partial z}\right)\cdot\left(A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}\right) \tag{B.20}$$

or

$$\boldsymbol{\nabla}\cdot\mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} = \lim_{\Delta V\to 0}\frac{\oint_{\mathtt{S}}\mathbf{A}\cdot d\mathbf{S}}{\Delta V} \tag{B.21}$$

where ΔV is the volume element.

In cylindrical polar coordinates, it is

$$\boldsymbol{\nabla}\cdot\mathbf{A} = \frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho A_\rho) + \frac{1}{\rho}\frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_z}{\partial z} \tag{B.22}$$

In spherical coordinates, it leads to

$$\boldsymbol{\nabla}\cdot\mathbf{A} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 A_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}(A_\theta\sin\theta) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial \phi} \tag{B.23}$$

Divergence of a cross product is zero:

$$\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{A}) = 0 \tag{B.24}$$

The Gauss divergence theorem is given by

$$\int_{\mathtt{V}}\boldsymbol{\nabla}\cdot\mathbf{A}d\mathtt{V} = \oint_{\mathtt{S}}\hat{\mathbf{n}}\cdot\mathbf{A}d\mathtt{S} \tag{B.25}$$

where $\hat{\mathbf{n}}$ is the unit vector.

© 2015 by Taylor & Francis Group, LLC

B.2.3 Laplacian

The divergence of the gradient of a scalar function is called the Laplacian, which is given in Cartesian coordinates as

$$\boldsymbol{\nabla}\cdot\boldsymbol{\nabla} f = \boldsymbol{\nabla}^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \tag{B.26}$$

In cylindrical polar coordinates, the Laplacian of a scalar f becomes

$$\boldsymbol{\nabla}^2 f = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2} \tag{B.27}$$

and in spherical coordinates, the Laplacian of a scalar f becomes

$$\boldsymbol{\nabla}^2 f = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial f}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial \phi^2} \tag{B.28}$$

B.2.4 Curl

The curl of a vector field, denoted by curl $(\mathbf{A})$ or $\boldsymbol{\nabla}\times\mathbf{A}$, is defined as the vector field having magnitude equal to the maximum circulation at each point and being oriented perpendicularly to this plane of circulation for each point. The differentiation of a vector is a vector; the curl operation can produce that. In the case of $\boldsymbol{\nabla}\times\mathbf{A} = 0$, the field is said to be an irrotational field. In Cartesian coordinates, the curl is defined by

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{A} &= \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)\hat{\mathbf{i}} + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)\hat{\mathbf{j}} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\hat{\mathbf{k}} \\ &= \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix}\end{aligned} \tag{B.29}$$

In cylindrical coordinates, it is written as

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{A} &= \left(\frac{1}{\rho}\frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z}\right)\hat{\boldsymbol{\rho}} + \left(\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho}\right)\hat{\boldsymbol{\phi}} \\ &\quad + \left(\frac{1}{\rho}\frac{\partial(rA_\phi}{\partial \rho} - \frac{1}{\rho}\frac{\partial A_\rho}{\partial \phi}\right)\hat{\mathbf{k}}\end{aligned} \tag{B.30}$$

and in spherical coordinates, it is written as

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{A} &= \frac{\hat{\mathbf{r}}}{r\sin\theta}\left(\frac{\partial}{\partial \theta}(A_\phi\sin\theta) - \frac{\partial A_\theta}{\partial \phi}\right) + \frac{\hat{\boldsymbol{\theta}}}{r}\left(\frac{1}{\sin\theta}\frac{\partial A_r}{\partial \phi} - \frac{\partial}{\partial r}(rA_\phi)\right) \\ &\quad + \frac{\hat{\boldsymbol{\phi}}}{r}\left(\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial \theta}\right)\end{aligned} \tag{B.31}$$

© 2015 by Taylor & Francis Group, LLC

To note, the curl of a gradient vanishes, that is,

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}A = 0 \tag{B.32}$$

Triple vector identity is

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\mathbf{A} = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A}) - \boldsymbol{\nabla}^2\mathbf{A} \tag{B.33}$$

The LHS of equation (B.33) is the curl of the curl of a vector field. The first term on the RHS is known as the gradient of the divergence and the second term is the Laplacian:

$$\mathbf{A}\cdot(\boldsymbol{\nabla}\times\mathbf{B}) - \mathbf{B}\cdot(\boldsymbol{\nabla}\times\mathbf{A}) = -\boldsymbol{\nabla}\cdot(\mathbf{A}\times\mathbf{B}) \tag{B.34}$$

B.3 Stokes Theorem

According to Stokes theorem, the line integral of a vector function around a closed contour is equal to the integral of the normal component of the curl of that vector function over any surface having the contour ($\mathtt{C}$) as its boundary edge, that is,

$$\int_{\mathtt{S}} (\boldsymbol{\nabla}\times\mathbf{A})\cdot\hat{\mathbf{n}}\,d\mathtt{S} = \oint_{\mathtt{C}} \mathbf{A}\cdot d\mathbf{l} \tag{B.35}$$

where ∇ represents a vector differential operator, $\hat{\mathbf{n}}$ an outward-drawn unit normal to the element of surface area $d\mathtt{S}$, and the components of the vector field $\mathbf{A}$ are continuous and have first partial derivatives within and on the boundary of $\mathtt{C}$.

By incorporating the limit of $\oint \mathbf{E}\cdot d\mathbf{l}$ as the area $\Delta\mathtt{S}$ approaches zero, on using the vector definition for the curl, we obtain

$$\mathbf{n}\cdot\boldsymbol{\nabla}\times\mathbf{A}(\mathbf{r}) \equiv \lim_{\Delta\mathtt{S}\to 0}\left[\frac{1}{\Delta\mathtt{S}}\oint_{\mathtt{C}} \mathbf{A}\cdot d\mathbf{l}\right] \tag{B.36}$$

B.4 Green's Theorem

Green's theorem is a vector identity which is equivalent to the curl theorem in the plane. It provides the relationship between a line integral around a closed curve and a double integral over the plane region. Let us consider that ϕ and ψ are two complex-valued functions of position, and $\mathtt{S}$ is a closed surface

© 2015 by Taylor & Francis Group, LLC

surrounding a volume V. If ϕ, ψ, and their first and second partial derivatives are single-valued and continuous within and on S, we can have

$$\int_{\mathtt{V}} \left(\phi\nabla^2\psi - \psi\nabla^2\phi\right) d\mathtt{V} = \int_{\mathtt{S}} \left(\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\right) d\mathtt{S} \tag{B.37}$$

in which $\frac{\partial}{\partial n}$ denotes a differentiation in the inward normal direction at each point on S.

© 2015 by Taylor & Francis Group, LLC

Bibliography

[1] C. Gittelson, www.mathematik.uni-dortmund.de/tdohnal/SOLIT.../kdv_handout.pdf.

[2] V. M. Nakariakov, Solar MHD Notes. University of Warwick, 2002.

[3] A. Satya Narayanan, Studies on Internal Gravity Waves in Shear Flows and Their Stability Characteristics, Ph.D. thesis, Indian Institute of Science, Bangalore, India, 1982.

[4] V. S. Pandey, B. N. Dwivedi, New Astron., 12, 182, 2006.

[5] J. Ireland, E. R. Priest, Solar Phys., 173, 31, 1997.

[6] R. Erdelyi, Solar Phys., 180: 213, 1998.

[7] K. Somasundaram, C. Uberoi, Solar Phys., 81, 19, 1982.

[8] A. Satya Narayanan, V. S. Pandey, P. Venkatakrishnan, ESPM - 12, Freiburg, Germany, 2008.

[9] R. Erdelyi, E. Varga, M. Zetenyi, 9th ESPM, Florence, Italy, 269, 1999.

[10] E. Varga, R. Erdelyi, ESA SP 464, 255, 2001.

[11] B. Roberts, Magnetohydrodynamic Surface Waves, Mechanisms of Chromospheric and Coronal Heating, Eds. P. Ulmschneider, E. R. Priest and R. Rosner, Springer-Verlag, Heidelberg, 2001.

[12] P. M. Edwin, B. Roberts, Solar Phys., 88, 179, 1983.

[13] A. Satya Narayanan, C. Kathiravan, R. Ramesh, I A U Symposium, 257, 563, 2009.

[14] A. Satya Narayanan, R. Ramesh, C. Kathiravan, E. Ebenezer, ASP Conf. Ser. 325, 295, 2004.

[15] R. Erdelyi, Physics of Sun and Its Atmosphere, Eds. B. N. Dwivedi & U. Narain, 61, World Scientific, Singapore, 2008.

[16] B. K. Carter, R. Erdelyi, Astron. & Astrophys., 481, 239, 2008.

[17] R. Erdelyi, V. Fedun, Solar Phys., 238, 41, 2006.

© 2015 by Taylor & Francis Group, LLC

[18] V. M. Nakariakov, L. Ofman, Astron. & Astrophys., 372, L53, 2001.

[19] C. T. Russel, M. M. Hoppe, W. A. Livesay, Nature, 296, 45, 1982.

[20] K. Miyomoto, Fundamentals of Plasma Physics and Controlled Fusion, Lecture Notes, www.nifs.ac.jp/report/NIFS-PROC-88.pdf, 2000.

[21] T. S. Bastian, Notes on Electromagnetic Waves in a Plasma, hesperia.gsfc.nasa.gov/summerschool/bastian/Notes-Waves.pdf, 2005.

[22] E. E. Gossard, J. Atmos. Sci., 31, 483, 1974.

[23] L. Gramer, Geophys. Fluid Dyn. Lect., II, 1, 2007.

[24] L. H. Lyu, Nonlinear Space Plasma Physics (I) docsfiles.com/pdf-chapter-5-instabilities.html, 2005.

[25] H. Hertz, Annal. der Physik, 267, 983, 1887.

[26] J. Strutt, Phil. Mag., 41, 447, 1871.

[27] L. Brillouin, Annal. des Phys., 17, 88, 1922.

[28] A. H. Compton, Phys. Rev., 21, 483, 1923.

[29] M. Planck, Annal. der Physik., 4, 553, 1901.

[30] C. H. Townes, A. Schawlow, Phys. Rev., 112, 1940, 1958.

[31] C. J. Doppler, Abhandlungen der Königl. Böhm. Gesellschaft der Wissenschaften, 2, 465, 1842.

[32] S. Morel, S. K. Saha, *21st Century Astrophysics, Eds. S. K. Saha, & V. K. Rastogi*, 2005, Anita Publications, New Delhi.

[33] A. Labeyrie, Astrophys. J., 196, L71, 1975.

[34] T. Young, Phil. Trans. Roy. Soc., London, XCII, 12, 387, 1802.

[35] A. A. Michelson, E. W. Morley, Am. J. Sci., 34, 333, 1887.

[36] G. Kirchhoff, berl. Ber., 641, 1882.

[37] A. Babinet, Compt. Rend., 4, 638, 1837.

[38] L. N. Hazra, Private communication, 2012.

[39] C. A. Coulomb, Histoire de lÁcadémie Roy. des Sci., 67, 1786.

[40] E. Hall, Am. J. Math., 2, 287, 1879.

[41] H. C. Oersted, Annal. Phil., 16, 273, 1820.

[42] J. J. Thompson, Phil. Mag., 5, 293, 1897.

© 2015 by Taylor & Francis Group, LLC

[43] J. H. Poynting, Phil. Trans. Roy. Soc. Lond., 175, 343, 1884.

[44] N. D. Lang, W. Kohn, Phys. Rev. B, 3, 1215, 1971.

[45] R. F. Donnelly, J. Geophys. Res., 72, 5247, 1967.

[46] C. O. Hines, Canad. J. Phys., 38, 1441, 1960.

[47] C. O. Hines, J. Geophys. Res., 70, 177, 1965.

[48] J. Testud, J. Atmos. Terr. Phys., 32, 1793, 1970.

[49] S. K. Saha, Ind. J. Phys., 75B, 285, 2001.

[50] A. K. Sen, B. Saha, S. K. Saha, S. K. Trehan, J. S. Sehra, R. N. Ghosh, J. Inst. Electron. Telecom. Engrs., 24, 320, 1978.

[51] S. K. Saha, Ph.D. Thesis, 'Studies on the Ionospheric Radio Wave Propagation Characteristics of Standard Frequency Signals in the HF Band at Calcutta, Calcutta University, 1983.

[52] M. S. Ruderman, J. V. Hollweg, M. Goossens, Physics of Plasmas 4, 75, 1997.

[53] P. S. Joarder, V. M. Nakariakov, B. Roberts, Solar Phys., 176, 285, 1997.

[54] M. J. Aschwanden, J. C. Brown, E. P. Kontar, Solar Phys., 210, 383, 2002.

[55] A. Satya Narayanan, R. Ramesh, Magnetohydrodynamic Waves in the Solar Corona, in *21st Century Astrophysics, Eds. S. K., Saha, V. K., Rastogi*, 213, Anita Publications, New Delhi, 2001.

[56] B. Roberts, P. M. Edwin, A. O. Benz, Astrophys. J., 279, 857, 1984.

[57] M. S. Ruderman, E. N. Pelinovsky, B. Roberts, N. S. Petrukhin, Physics of Plasmas, 8, 2682, 2001.

[58] V. M. Nakariakov, E. Verwichte, Living Rev. Solar Phys. 2, 3, 2005.

[59] G. E. Moreton, Astron. J. 65, 494, 1960.

[60] G. E. Moreton, H. Ramsey, Pub. Astron. Soc. Pacific., 72, 354, 1960.

[61] Y. Uchida, Solar Phys. 39, 431, 1974.

[62] W. M. Neupert, Astrophys. J., 344, 504, 1989.

[63] D. Moses et al., Solar Phys., 175, 571, 1997.

[64] D. Biskamp, Nuclear Fusion, 13, 719, 1973.

[65] H. Bateman, Monthly Weather Rev., 43, 163, 1915.

© 2015 by Taylor & Francis Group, LLC

[66] J. M. Burgers, Proc. Roy. Neth. Acad. Sci., 43, 2, 1940.

[67] E. Hopf, Commun. Pure & Appl. Math., 3, 201, 1950.

[68] J. D. Cole, Quart. Appl. Math., 9, 225, 1951.

[69] M. J. Lighthill, Surveys in Mechanics., Eds. G. K. Batchelor & R. M. Davies, 250, 1956.

[70] G. I. Taylor, Proc. Roy. Soc. A, 84, 371, 1910.

[71] J. C. Nimmo, D. G. Crighton, Proc. Roy. Soc. A 384, 381, 1982.

[72] W. Heisenberg, Zeitschrift für Physik., 43, 172, 1927.

[73] A. Einstein, Annal. der Physik, 17, 132, 1905a.

[74] N. Bohr, Phil. Mag., 26, 151, 1913.

[75] E. Schrödinger, Phys. Rev., 28, 1049, 1926.

[76] E. Wolf, Proc. Roy. Soc. A, 225, 96, 1954.

[77] A. Einstein, Annal. der Physik, 17, 891, 1905b.

[78] A. Einstein, Annal. der Physik, 17, 547, 1905c.

[79] A. Einstein, Annal. der Physik, 18, 639, 1905d.

[80] P. A. M. Dirac, Proc. Roy. Soc. Lond., Ser. A, 117, 610, 1928.

[81] C. D. Anderson, Phys. Rev., 43, 491, 1933.

[82] G. G. Stokes, Trans. Cambridge Phil. Soc., 9, 399, 1926.

[83] H. Poincaré, Theorie Mathématique de la Lumiére, II, Chap 12, 1892.

[84] V. Piirola, Astron. & Astrophys, 27, 382, 1973.

[85] E. H. Land, J. Opt. Soc. A., 41, 957, 1951.

[86] R. C. J. Jones, Opt. Soc. A., 31, 488, 1941.

[87] H. Mueller, J. Opt. Soc. Am., 37, 110, 1947.

[88] A. T. Young, Astron. J., 72, 747, 1967.

[89] W. Wasow, Ann. Math., 53, 350, 1950.

[90] A. L. Rabenstein, Arch Ration Mech Anal., 1, 418, 1958.

[91] A. Hillier, T. Berger, H. Isobe, K. Shibata, Astrophys. J. 746, 110, 2012.

[92] S. Ali, Z. Ahmed, A. M. Mirza, I. Ahmed, Phys. Lett. A 373, 2940, 2009.

© 2015 by Taylor & Francis Group, LLC

[93] C. Foullon, E. Verwichte, V. M. Nakariakov, K. Nykyri, C. J. Farrugia, Astrophys. J., 729, L8, 2012.

[94] L. Ofman, B. J. Thompson, arXiv: 1101 4249v2, 20011.

[95] F. Malara, L. Primavera, P. Veltri, Nonlin. Process. Geophys. 8, 159, 2001.

[96] L. D. Zanna, M. Velli, P. Londrillo, Astron. & Astrophys., 367, 705, 2001.

[97] D. A. Gurnett, J. E. Maggs, D. L. Gallagher, W. S. Kurth, F. L. Scarf, J. Geophys. Res., 86, 8833, 1981.

[98] J. C. Weatherall, M. V. Goldman, D. R. Nicholson, Astrophys. J., 246, 306, 1981.

[99] M. Ablowitz and H. Segur. *Solitons and Inverse Scattering Transform*. SIAM, Philadelphia, 1981.

[100] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. McGraw-Hill, New York, 1972.

[101] G. Arfken. *Mathematical Methods for Physicists*. Academic Press, Orlando, 2012.

[102] W. Baumjohann and R. A. Treumann. *Basic Space Plasma Physics*. Imperial College Press, London, UK, 1996.

[103] P. M. Bellan. *Fundamentals of Plasma Physics*. Cambridge University Press, UK, 2006.

[104] J. M. Bennett. *Handbook of Optics*. McGraw-Hill, New York, 1994.

[105] M. Born and E. Wolf. *Principle of Optics*. Pergamon Press, 1984.

[106] R. Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill, New York, 1965.

[107] L. Brillouin. *Wave Propagation in Periodic Structures: Electric Filters and Crystal Lattices*. Dover, New York, 2003.

[108] C. Brosseau. *Fundamentals of Polarized Light: A Statistical Optics Approach*. John Willey and Sons, New York, 1998.

[109] B. F. Burke and F. Graham-Smith. *An Introduction to Radio Astronomy*. Cambridge University Press, UK, 1998.

[110] J. Byrnes. *Unexploded Ordnance Detection and Mitigation*. Springer, 2009.

© 2015 by Taylor & Francis Group, LLC

[111] R. A. Cairns. *Plasma Physics*. Blackie, Glasgow, UK, 1985.

[112] S. Chandrasekhar. *Radiative Transfer*. Dover Publications, New York, 1960.

[113] S. Chandrasekhar. *Hydrodynamic and Hydromagnetic Stability*. Oxford University Press, UK, 1961.

[114] F. F. Chen. *Introduction to Plasma Physics*. Plenum, New York, 1977.

[115] E. Collett. *Polarized Light: Fundamentals and Applications*. Marcel Dekkar, New York, 1993.

[116] F. S. Crawford. *Waves*. McGraw-Hill, New York, 1968.

[117] R. C. Davidson. *Nonlinear Plasma Theory*. Academic, New York, 1972.

[118] R. O. Dendy. *Plasma Dynamics*. Oxford University Press, 1990.

[119] O. K. Ersoy. *Diffraction, Fourier Optics and Imaging*. Wiley and Sons, New York, 2006.

[120] R. Fitzpatrick. *Oscillations and Waves*. CRC Press, Taylor and Francis, USA, 2013.

[121] E. B. Fomalont and M. C. H. Wright. *Galactic and Extra-galactic Radio Astronomy*. Addison-Wesley, USA, 1974.

[122] M. Francon. *Optical Interferometry*. Academic Press, New York, 1966.

[123] A. P. French. *Vibrations and Waves*. W. W. Norton and Co., New York, 1971.

[124] D. Goldstein. *Polarized Light*. Dekker, New York, 2003.

[125] H. Goldstein. *Classical Mechanics*. Addison-Wesley, MA, 1980.

[126] R. J. Goldston and P. H. Rutherford. *Introduction to Plasma Physics*. Institute of Physics Publishing, Bristol, UK, 1995.

[127] J. W. Goodman. *Introduction to Fourier Optics*. John Wiley and Sons, New York, 1968.

[128] E. E. Gossard and W. H. Hooke. *Waves in the Atmosphere*. Elsevier, 1975.

[129] I. S. Gradshteyn and I. M. Ryzhik. *Tables of Integrals, Series, and Products*. Academic Press, San Diego, 2000.

[130] D. J. Griffiths. *Introduction to Electrodynamics*. Prentice-Hall, London, 1999.

© 2015 by Taylor & Francis Group, LLC

[131] D. Halliday, R. Rasnick, and J. Walkar. *Fundamentals of Physics*. John Wiley and Sons, New York, 2001.

[132] A. Hirose and K. E. Lonngren. *Introduction to Wave Phenomena*. John Wiley and Sons, New York, 1985.

[133] J. D. Jackson. *Classical Electrodynamics*. Wiley, New York, 1999.

[134] E.C. Jordan and K. G. Balmain. *Electromagnetic Waves and Radiating Systems*. Prentice-Hall Inc., New Jersey, 1968.

[135] S. G. Krantz. *Handbook of Complex Variables*. Boston, MA, 1999.

[136] J. D. Krauss. *Radio Astronomy*. McGraw Hill, New York, 1966.

[137] A. Labeyrie, S. G. Lipson, and P. Nisenson. *Introduction to Stellar Optical Interferometry*. Cambridge University Press, UK, 2006.

[138] G. L. Lamb. *Elements of Soliton Theory*. Wiley, New York, 1980.

[139] H. Lamb. *Hydrodynamics*. Cambridge University Press, UK, 1932.

[140] L. D. Landau and E. M. Lifshitz. *Fluid Mechanics*. Butterworth Heinmann, 1997.

[141] W. Lauterborn and T. Kurz. *Coherent Optics*. Springer-Verlag, Berlin, 2003.

[142] M. J. Lighthill. *Waves in Fluids*. Cambridge University Press, UK, 2001.

[143] I. G. Main. *Vibrations and Waves in Physics*. Cambridge University Press, U K, 1984.

[144] L. Mandel and E. Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, UK, 1995.

[145] M. Mansfield and C. O'Sullivan. *Understanding Physics*. Wiley and Sons, UK, 2011.

[146] J. M. Mariotti. *Diffraction-Limited Imaging with Very Large Telescopes, Editors, D. M. Alloin and J. -M. Mariotti*. Kluwer Academic Publishers, 1988.

[147] J. C. Maxwell. *A Treatise on Electricity and Magnetism*. Oxford University Press, UK, 1873.

[148] A. Satya Narayanan. *An Introduction to Waves and Oscillations in the Sun*. Springer, New York, 2013.

[149] A. H. Nayfeh. *Perturbation Methods*. John Wiley, 2004.

© 2015 by Taylor & Francis Group, LLC

[150] A. Papoulis. *Systems and Transforms with Applications in Optics.* McGraw-Hill, New York, 1968.

[151] G. K. Parks. *Physics of Space Plasmas.* Addison-Welsley Publication, 1991.

[152] J. Pedlosky. *Geophysical Fluid Dynamics*. Springer-Verlag, Berlin, 1979.

[153] F. L. Pedrotti and L. S. Pedrotti. *Introduction to Optics.* Prentice Hall Inc., New Jersey, 1987.

[154] E. R. Priest. *Solar System Magnetic Fields.* D. Reidel Publishers, Holland, 1982.

[155] E. M. Purcell. *Electricity and Magnetism.* McGraw-Hill, New York, 1985.

[156] R. Leighton R. Feynman and M. Sands. *The Feynman Lectures on Physics*. Addison-Wesley, USA, 1964.

[157] J. A. Ratcliffe. *The Magneto-Ionic Theory and Its Applications to the Ionosphere.* Cambridge University Press, 1959.

[158] H. W. Reddick and F. H. Miller. *Advanced Mathematics for Engineers.* Willey and Sons Inc, New York, 1970.

[159] P. L. Sachdev. *Nonlinear Diffusive Waves*. Cambridge University Press, UK, 1987.

[160] S. K. Saha. *Diffraction-limited Imaging with Large and Moderate Telescopes.* World-Scientific, New Jersey, 2007.

[161] S. K. Saha. *Aperture Synthesis: Methods and Applicatios to Optical Astronomy.* Springer, New York, 2010.

[162] S. K. Saha. *High Resolution Imaging: Detectors and Applications*. Pan Stanford Publishing, Singapore, 2015.

[163] K. Serbowski. *Planets, Stars, and Nabulae Studied with Photopolarimetry.* University of Arizona Press, Tucson, 1947.

[164] B. K. Shivamoggi. *Theoretical Fluid Dynamics.* Martinus Nijhoff, 1985.

[165] W. T. Silfvast. *Laser Fundamentals*. Cambridge University Press, UK, 2004.

[166] G. L. Smith. *An Introduction to Classical Electromagnetic Radiation.* Cambridge University Press, UK, 1997.

[167] A. Sommerfeld. *Optics, Lectures on Theoretical Physics.* Academic Press, New York, 1952.

© 2015 by Taylor & Francis Group, LLC

[168] T. H. Stix. *The Theory of Plasma Waves*. 1st Ed. McGraw-Hill, New York, 1962.

[169] P. A. Sturrock. *Plasma Waves*. Cambridge University Press, UK, 1994.

[170] D. G. Swanson. *Plasma Waves*. Institute of Physics Publishing, Bristol and Philadelphia, 2003.

[171] R. A. Thompson, J. M. Moran, and G. W. Swenson. *Interferometry and Synthesis in Radio Astronomy*. John Wiley and Sons, Inc, New York, 2001.

[172] H. C. van de Hulst. *Light Scattering by Small Particles*. Dover Publications, New York, 1957.

[173] R. A. Waldron. *Waves and Oscillations*. D. Van Nostraxd Company, INC., Canada, 1964.

[174] G. B. Whitham. *Linear and Nonlinear Waves*. Wiley Inter Science, 1999.

© 2015 by Taylor & Francis Group, LLC

Index

KdV, 230, 312

Aberration, 53
Acceleration, 189, 190, 204
Acoustic waves, 240, 248
Acoustic-gravity waves, 249
Airy equation, 436
Alfven speed, 239, 241, 249, 255
Alfven waves, 235, 240, 241, 257
Ambient shear, 457
Ampère−Maxwell law, 115, 117, 118, 121, 123
Ampère's law, 102, 104, 110, 115, 116
Amplitude, 160, 316
Angular momentum, 124
Angular velocity, 236
Anisotropic, 416
Antenna, 140
  Dipole, 142, 152
  Dish, 152
  Gain, 151
  Monopole, 150
  Radiation pattern, 142
Asymmetric mode, 289
Asymptotic behavior, 318
Asymptotic form, 349
Asymptotic solutions, 436, 437
Atom, 2
Atomic clock, 156
Aurora, 155
Auroral, 422

Backward waves, 292
Bandwidth, 373
Benjamin–Ono, 268, 307
Bessel function, 63
Bessel functions, 296
Biot−Savart law, 102, 104, 110, 144, 147
Black body, 14
Bohm–Gross relation, 430
Boltzmann gas, 332
Boundary conditions, 196, 229
Boussinesq, 188, 209
Boussinesq approximation, 249, 251
Bow shock, 327
Brunt–Vaisala, 194, 227, 247
Bulk waves, 220
Burger's equation, 268, 308, 312

Capacitor, 97, 111, 116, 142
Capillary, 198
Cauchy part, 308
Cavity, 164
Chain rule, 212
Characteristic frequency, 418
Characteristic scale, 258
Charge, 74, 84
  Distribution, 80
Charge cloud, 417
Charge density, 80
  Linear, 80
  Surface, 80
  Volume, 80
Charge flux, 483
Charge neutrality, 422
Chromosphere, 290, 292
Circularly polarized, 427, 428
Cnoidal wave, 184
Coefficient matrix, 215
Coefficient of diffusion, 258
Coherence, 34
  Coherence length, 34, 375
  Coherence time, 35, 372

© 2015 by Taylor & Francis Group, LLC

Partial coherence, 34
Spatial, 35
Temporal, 35
Cole–Hopf transformation, 343
Collision, 181
Collisional damping, 438
Collisionless plasmas, 328
Collisionless shock, 328, 330
Complex amplitudes, 217
Complex conjugate, 201
Compressibility, 270, 272
Compressional viscosity, 255
Conservation, 187, 223
Continuity equation, 121, 207
Coriolis forces, 452
Corona, 293
Coronal, 308
Coronal mass ejections, 326
Coronal plasma, 263
Corpuscular theory, 3
Coulomb collisions, 328, 329
Coulomb's law, 74, 91
Critical levels, 457
Cross helicity, 477
Cross product, 497
Current, 74, 91, 98
Alternating current, 110, 111
Conduction current density, 115
Convection current density, 116
Current density, 98, 104, 114–116
Direct current, 102, 110
Displacement current, 117, 144
Displacement current density, 114
Stationary current, 102
Steady current, 102
Curvilinear coordinate, 257
Cusp speeds, 307
Cut−off, 305
Cylinder, 267, 293

D'Alembert, 171
Damped, 165
Damping, 163, 251, 253, 261
Debye length, 342, 417, 449
Debye shielding, 417
Delta function, 198, 201
Density, 188, 189, 203
Detector, 33
Dielectric, 420
Dielectric constant, 32, 153
Dielectric tensor, 417
Differential operators, 447
Diffraction, 43
Airy pattern, 60
Debye approximation, 52
Fresnel approximation, 53
Fresnel integral, 54
Fresnel zone, 44
Fresnel−Kirchhoff formula, 50
Huygens principle, 43
Huygens−Fresnel principle, 44
Paraxial approximation, 52
Rayleigh−Sommerfeld integral, 51
Diffusion, 174
Dipole, 78, 156
Electric, 79
Hertzian, 143
Magnetic, 99
Torque, 79
Dirac quantization, 359
Discontinuity, 260, 263
Dispersion, 23
Dispersion relation, 194, 235, 248, 272, 279, 286, 290, 291, 293, 301, 303, 304, 310, 421, 425, 447
Dispersive, 328
Displacement, 160, 165, 195, 202, 208, 229
Dissipative effects, 311
Dissipative layer, 261, 263–265
Distribution function, 335, 339, 416
Divergence, 188
Divergence-free, 472
Doppler broadening, 31
Doppler effect, 30, 373
Doppler-shift, 6, 30, 153

© 2015 by Taylor & Francis Group, LLC

Doppler-shifted, 250
Doppler-shifted frequency, 472
Dot product, 496
Drift waves, 432
Dynamical system, 452

Earth, 156
Eccentricity, 389
Eigensolution, 455, 459
Eigenvalue, 454
Eigenvalue problem, 214, 224
Electric current, 236
Electric field, 77, 81, 92, 104, 107, 146, 416
Electric potential, 79, 90, 92, 104, 417
Electricity, 153
Electromagnetic, 187
Electromagnetic radiation, 2
Electromagnetic wave, 417
Electromotive force, 91, 107, 108, 115
Electron, 2, 416
Electron-volt, 3
Electrons, 153
Electrostatic potential, 341
Electrostatics, 74, 90
Elliptic functions, 184
Ellipticity, 378, 389
Energy, 75
  Kinetic, 100
  Potential, 76
Energy flux, 312
Enthalpy, 314
Entropy, 314, 331
Equilibrium, 168, 193, 206
Ergodicity, 32
Euler equations, 192
Euler formula, 361
Evanescent, 270, 281, 295
Exponentially diverging, 259
Extraordinary wave, 426

Faraday rotation, 428
Faraday's law, 91, 107, 108, 114
Fermat law, 354
Filaments, 326
Flare, 302
Flares, 326
Flute, 265
Flute instability, 478, 482
Flux, 81
  Density, 83
  Electric, 81, 85
  Magnetic, 82
Flux tubes, 262, 292
Force, 74
  Magnetic, 106
Fourier theorem, 21, 360
Fourier transform, 18, 26, 56, 65, 139, 195, 200, 360
  Convolution, 54, 363
  Parseval theorem, 140, 363, 366, 372
  Power spectrum, 363
  Power theorem, 140
  Rayleigh theorem, 363
Free oscillations, 212, 223
Free surface, 192, 210
Frequencies, 269
Frequency, 2, 3, 10, 137, 156, 194, 217, 236, 262
  Angular, 111, 133
  Electrical, 110, 144
  Gamma ray, 3
  HF, 5, 156
  Infrared, 4
  Interval, 366, 373
  LF, 5, 156
  MF, 5
  Microwave, 5
  Optical, 356
  Optical band, 4
  Radio, 5
  Resonant, 113
  UV ray, 4
  VHF, 5
  VLF, 5, 156
  X-ray, 4
Frequency shift, 447
Frictionless, 160

© 2015 by Taylor & Francis Group, LLC

Gas dynamics, 311
Gas pressure, 235, 245, 261
Gauss' law, 81, 94, 95
    Electric field, 118
    Magnetic field, 119
Gaussian, 178
Gaussian surface, 83, 89
General solution, 319, 331
Geometric optics, 432
Geostrophic, 192, 223
Granules, 292, 294
Grating, 65
Gravity, 163
Gravity Waves, 198
Green's theorem, 45, 501
Group velocity, 194, 203, 448
Growth rate, 470, 484
Gyro motion, 427

Hall effect, 106
Hankel function, 438
Harmonic, 160, 169
Harmonic oscillator, 418
Harmonic wave, 201
Heat equation, 344
Heat pulse, 184
Helix, 419
Helmholtz, 172
Helmholtz equation, 45
Hilbert transform, 308, 361
Homogeneous, 195, 204, 215
Howard's semi-circle, 456
Huygens principle, 3, 39
Huygens principle, 437
Hybrid frequency, 426
Hybrid oscillations, 421
Hybrid resonance layer, 438
Hydrostatic, 193, 205, 228

Impedance, 111
Incompressible, 187
Inductance, 108
    Mutual inductance, 109
    Self-inductance, 108
Inductor, 110, 111, 142
Inertia, 163
Inertial, 221
Inertial waves, 188
Initial value problem, 320
Instabilities, 331, 335
Instantaneous, 161
Integral equation, 323
Intensity, 17, 32, 128, 136
Interference, 34
    Michelson interferometer, 41
    Visibility, 38
    Young's experiment, 39, 127
    Young's principle, 355
Interferometry, 35
Internal energy, 189, 314
Internal gravity waves, 193
Internal stresses, 452
Internal-Alfven-Gravity waves, 249
Inverse square law, 17
Ion, 416, 422
Ion-acoustic waves, 341
Ionosphere, 153, 423
    D region, 153
    E region, 153
    F region, 153
    F_1 region, 153
    F_2 region, 153
    SFD, 155
    SID, 154
    TID, 153
ions, 153
Isentropic energy, 297
Isolines, 192

Jacobian, 192, 257
Joule heat, 130
Joule heating, 155

Kadomstev−Petviashvili, 185
KdV, 229
Kelvin, 188, 216, 452
Kinematic, 204
Kink, 265
Kink speed, 295, 304
Kirchhoff law, 112

© 2015 by Taylor & Francis Group, LLC

Klein−Gordon, 188
Klein−Gordon equation, 299

Lagrangian, 261
Lagrangian pressure, 300
Lamb, 218, 220
Landau damping, 438
Laplace equation, 196
Laplace transform, 20
Laplacian operator
  Cartesian coordinate, 498
  Cylindrical polar coordinate, 498
  Spherical polar coordinate, 498
Laser, 25
Lenz law, 107
Light source, 14
  Luminescent, 14
  Thermal, 14
Linearize, 190, 191
Long wavelength, 285, 288, 304
Long waves, 197, 199, 229
Lorentz factor, 358
Lorentz force, 106, 124, 419
Lorentz law, 129

Mach angle, 336
Mach number, 312, 315, 325
Magnetic diffusion, 264
Magnetic field, 107, 146
Magnetic fields, 235
Magnetic induction, 263, 265
Magnetic permeability, 294
Magnetic potential, 104, 145
Magnetic storm, 155
Magnetically structured, 268
Magnetization, 123
Magneto acoustic-gravity (MAG), 282
Magneto acoustic-gravity waves, 249
Magneto hydrostatics, 282
Magneto sonic, 269
Magnetomotive force, 117
Magnetosphere, 153, 427
Mass flux, 312, 324
Matched asymptotic expansion, 264
Maximum amplitude, 164
Maxwell equations, 354, 422
Maxwell−Boltzmann distribution, 31
Maxwell's equations, 2, 45, 114, 126, 133, 134, 153, 494
  First law, 114
  Fourth law, 119
  Second law, 115
  Sinusoidal fields, 120
  Third law, 118, 123
Mean free path, 328, 331
Method of characteristics, 319
Miles–Howard theorem, 461
Mode conversion, 432, 433
Molecular viscosity, 189
Momentum density, 442
Momentum equation, 189, 210, 229
Monoatomic, 325
Moreton wave, 326

Navier–Stokes equations, 431
Nonconservative, 189
Nonlinear, 263
Nonlinear couplings, 438
Nonlinear cubic, 184
Nonparallel propagation, 270, 277
Nonuniform media, 267
Normal modes, 263
Number density, 338

Ohm law, 99, 150
Ohmic dissipation, 250
Optical fiber, 184
Optical path difference, 41
Optics, 354
  Geometrical optics, 354
  Quantum optics, 355
  Statistical optics, 357
  Wave optics, 355
Ordinary wave, 424
Orthogonality, 174
Orthogonality condition, 258
Oscillating loop, 302
Oscillation, 9, 15, 26, 29, 111, 137, 159, 378

© 2015 by Taylor & Francis Group, LLC

Damped, 113
Nonharmonic, 368
Tidal, 155
Undamped, 113
Oscillator, 162

P-mode, 290
Parameters, 304
Parametric decay, 478
Permeability, 32, 101, 115, 236
Permittivity, 32, 75, 115
Perturbations, 193, 451, 487
Phase mixing, 255
Phase velocity, 197, 203, 207, 225, 417
Phonon, 8
Photon, 2, 356
Piecewise, 348
Piston effect, 327
Plane wave, 207, 212, 222
Plasma, 25, 153, 415, 423, 436
Poincaré, 188
Poisson equation, 95, 105, 341
Polarization, 149, 377
Analyzer, 391, 400, 410
Birefringence, 390
Brewster law, 391
Compensator, 392
Degree of depolarization, 387
Degree of polarization, 378, 387
Depolarizer, 389
Diattenuator, 390
Dichroism, 390
Eigen-polarization, 390
Half-wave plate, 393, 399, 407
Jones matrix, 395, 401
Malus law, 398
Mueller matrix, 401
Neutral density filter, 391, 398
Phase-shifter, 392, 404
Poincaré sphere, 386
Polarimeter, 409
Polarimetry, 409
Polarizer, 390, 407, 410
Quarter-wave plate, 392, 399, 406
Retarder, 392
Rotator, 393, 400
Spectropolarimetry, 409
Stokes parameter, 387, 401
Stokes parameters, 383
Stokes vector, 383
Wave plate, 392
Poles, 198
Polytropic gas, 312
Pondermotive force, 443
Potential temperature, 457
Power series, 342
Poynting flux, 339
Poynting theorem, 128
Poynting vector, 127, 148
Pressure, 188, 193
Probability, 357
Propagate, 199
Propagation velocity, 197, 203

Quantum mechanics, 357
Quasi parallel, 336
Quasi-linear, 347
Quasi-linear theory, 417, 449
Quasi-neutral, 417

Radioheliograph, 152
Rankine–Hugoniot, 314
Rayleigh, 220, 452, 460
Rectangular hyperbola, 324
Reductive perturbation, 308
Refractive index, 8, 11, 24, 153
Regular solution, 346
Resistor, 111
Resonances, 426, 432
Resonant absorption, 261
Resonant circuit, 110
Parallel, 113
Series, 112
Resonant term, 441
Resonator, 164
Reynolds number, 264, 348, 350
Riccati, 183

© 2015 by Taylor & Francis Group, LLC

Riccati equation, 344, 345
Richardson number, 460
Rigid surface, 204
Rotating frame, 188, 227

Sausage, 265
Sausage and kink instability, 478
Sausage mode, 285, 288, 299
Scalar, 495
Scalar product, 496
Scale-height, 249
Scaleheight, 237
Scattering, 7
    Brillouin scattering, 8
    Compton scattering, 8
    Elastic scattering, 7
    Inelastic scattering, 8
    Mie scattering, 8
    Raman scattering, 8
    Rayleigh scattering, 7
Schrödinger equation, 357
Schrodinger, 182
Schwarz inequality, 385
Sedov blast wave, 312
Self-similar, 316
Separable, 172
Series solution, 257
Shallow water, 199
Shallow water waves, 188, 210
Sharp gradients, 312
Shear flow, 284
Shear viscosity, 255
Shock adiabatic, 324
Shock front, 432
Shock strength, 321
Shock waves, 312, 341
Simple pendulum, 159
Single-hump, 348, 349
Single-valued, 311
Sink, 188
Sinusoidal, 176, 269, 288
Slab, 267, 270, 284, 308
Slow solitary, 306
Slow surface, 278, 287, 308
Slowly varying, 434
Snell law, 410, 411
Solar flare, 154
Solar wind, 154, 328
Solid angle, 81
Solitary, 231
Soliton, 185
Source, 188
Specific entropy, 189
Specific heat, 337
Specific heats, 315
Spectrograph, 6
Spectrometer, 6
Spectroscopy, 6, 409
Spectrum, 2, 201
    Absorption, 6
    Continuous, 5
    Emission line, 6
    Fraunhofer line, 6
Spherical shock, 316
Spicules, 292
Spring constant, 165
Square integrable, 456
Stability, 451, 478, 486, 489
Star, 5, 30, 61
Static magnetic field, 443
Stationary, 32
Steepening, 311, 328
Stokes, 452
Stokes parameter, 136
Stokes theorem, 107, 116, 501
Stokes vector, 129
Stratification, 193
Stratified, 283, 308
Stratified shear flows, 250
Stratified system, 452
Stream function, 225, 455
Stress tensors, 339
Stretched variables, 342
Strong shocks, 325
Sturm–Liouville problem, 456
Subcritical, 336
Subsonic, 269
Sun, 153
    Active, 154
    Sunspot, 154

© 2015 by Taylor & Francis Group, LLC

Sunspots, 293
Super critical, 336
Super thermal particles, 432
Superposition, 169
Surface, 248
Surface tension, 196–198
Surface Waves, 188
Sweeping-skirt, 326
Symmetric, 179, 201, 220
Symmetric mode, 289

Tangential discontinuities, 265, 311
Taylor expansion, 261
Taylor–Goldstein equation, 456
Taylor-Goldstein equation, 250
Thermal equilibrium, 422
Thermalization, 331
Time, 156
Transmission function, 56, 65
Trapping, 270
Traveling wave, 347, 348
Tube speed, 306
Twisted, 267
Type II bursts, 326

Uncertainty principle, 356, 374
Upward propagating, 258

Van Allen belts, 155
Vector, 495
  Addition, 496
  Component, 496
  Magnitude, 495
  Position vector, 496
  Subtraction, 496
  Unit vector, 495
Vector operator, 498
  Cartesian coordinate, 497
  Curl, 95, 104, 108, 131, 134, 139, 146, 500, 501
  Cylindrical polar coordinate, 498
  Del, 497
  Divergence, 93, 104, 119, 126, 127, 129, 133, 499
  Gradient, 91, 93, 96, 104, 499–501
  Laplacian, 15, 23, 95, 498, 500
  Spherical polar coordinate, 498
Vector product, 497
Velocity potential, 195, 201
Vibrations, 164
Visco resistive, 263
Viscous dissipation, 189
Visible matter, 415
Vlasov equation, 335
Vlasov equations, 440
Vorticity, 226, 232, 233

Wasow equation, 437
Wave, 9, 153
  Amplitude, 10
  Anti-node, 29
  Beat, 22
  Gravity wave, 155
  Group velocity, 21
  Longitudinal wave, 9
  Monochromatic, 25
  Node, 29
  Nonmonochromatic, 360
  Period, 12
  Phase, 12
  Phase velocity, 20
  Radio, 5, 153
  Sound wave, 9
  Standing wave, 29
  Stationary wave, 29
  Transverse wave, 9
  Traveling wave, 10
  Velocity, 12
  Water wave, 14
  Wave vector, 13
  Wavelength, 12
  Wavelet, 10, 43, 62
  Wavenumber, 13
Wave equation
  Electromagnetic, 132
Wave-particle duality, 357
Wavefront, 9, 327
  Plane, 10
Waveguide, 292, 293
Wavelength, 2, 3, 194, 235, 237, 244

© 2015 by Taylor & Francis Group, LLC

Wavelengths, 188
Wavenumber, 133, 196, 236, 253, 266, 272, 275, 305, 307
Waveparticle interaction, 439
Weak solution, 319, 323
Weak turbulence, 439
Whistler branch, 332
Whistler waves, 427
Work, 76

Young's modulus, 175

Zakharov–Shabat equation, 447

© 2015 by Taylor & Francis Group, LLC